CALENDAR

	2006	2007
JAN	Applied Business Research Conference 1/2-6: Lake Buena Vista, FL, USA Hawaii International Conference on System Sciences 1/4-7: Kauai, Hawaii ACM-SIAM Symposium on Discrete Algorithms 1/22-24: Miami, FL, USA Open University Winter Combinatorics Meeting 1/25: Milton Keynes, England, UK	Hawaii International Conference on System Sciences 1/3-6: Hilton Waikoloa Village Resort, Big Island, Hawaii, USA ACM-SIAM Symposium on Discrete Algorithms (SODA) 01/7-9: New Orleans, LA, USA SIENA workshops in Groningen 01/8-11, 16-20, 17-21: Groningen, The Netherlands 10th Annual Atmospheric Science Librarians International Meeting 01/17-19: San Antonio, TX, USA Network Centric Warfare 2007 01/22-25: Washington, DC, USA International Academy of Management and Business 01/28-31: Las Vegas, NV, USA
FEB	FRACTAL 2006 - Complexity and Fractals in Nature 2/12-15: Vienna, Austria The IASTED International Conference on Artificial Intelligence and Applications 2/13-16: Innsbruck, Austria	First International Conference of Aceh and Indian Ocean Studies 02/2: Banda Aceh, Indonesia SIAM Workshop on Combinatorial Scientific Computing (CSC07) 02/17-19: Costa Mesa, CA, USA Internatnl Conference on Environment: Survival & Sustainability 02/19-24: Nicosia, Cyprus Thinking Drinking II: From Problems to Solutions 02/26 – 28: Melbourne, Australia
MAR	ICIW: Information-Warfare & Security 3/15-16: U of Md. Eastern Shore, MD USA General Online Research (GOR06) 3/21-22: Ravensberger Park, Bielefeld, Germany	Asia-Pacific Academy of Management and Business 03/5-8: Singapore ICIW 2007: International Conference on i-Warfare and Security 03/8-9: Monterey, California, USA General Online Research (GOR07) 03/26-28: Universität Leipzig, Germany International Colloquium On Stochastic and Potential Analysis 03/26-29: Hammamet, Tunisia 4th Biennial Forum of N. American Ass'n of Fisheries Economists 03/27-30: Merida, Mexico
APR	Aveiro Workshop on Graph Spectra 4/10-12: Aveiro, Portugal European Meetings on Cybernetics and Systems Research (EMCSR) 4/18-21: Vienna, Austria 21st ACM Symposium on Applied Computing 4/23-27: Dijon, France International Sunbelt Social Networks Conference 4/25-30: Vancouver, BC, CA	International Coastal Symposium 04/16-20: Gold Coast, Queensland, Australia 3rd European Conference - Management, Leadership, Governance 04/19-20: University of Winchester, UK 15th International Conference on the Modelling, Monitoring and Management of Air Pollution 04/23-25: Algarve, Portugal SIAM International Conference on Data Mining (SDM07) 04/26-28: Minneapolis, MN, USA
MAY	International Congress on Medieval Studies: The Medieval Tradition of Natural Law 4/4-7: Kalamazoo, MI, USA 5th Intnl. Conference on Drugs and Young People 5/24-26: Randwick, NSW, Austalia	International Sunbelt Social Network Conference 05/1-6: Chandris Hotel, Corfu, Greece 6th International Triple Helix Conference on University, Industry and Government Linkages 05/16-18: National University of Singapore Fifth International Marine Bioinvasions Conference 05/21-24: Cambridge, MA, USA International Communication Association 05/24-28: San Francisco, CA, USA
JUN	Dynamics, Topology and Computations 6/4-10: Bedlewo, Poland International Communication Association 6/19-23: Dresden, Germany Conference on Stochastic Networks 6/19-24: University of Illinois, Urbana, IL USA International Conference on Topology and its Applications 6/23-26: Aegion, Greece	Network Centric Warfare Europe 2007 06/6-7: Prague, Czech Republic 21st Pacific Science Congress (21 PSC) 06/12-18: Okinawa, Japan 3rd nternational Conference on Business, Management, Economics 06/13-17: Izmir,Turkey 13th International Symposium on Society & Resource Management 06/18-21: Halifax, Nova Scotia, Canada

CALENDAR

2006	2007	
Second Oceanic Conference on International Studies 7/5-7: University of Melbourne, Melbourne, Australia Talcott Parsons: Reassessing his contribution to the social sciences 7/6-8: Manchester, UK Sociolinguistics Symposium 16 7/6-8: Limerick, Ireland IEEE World Congress on Computational Intelligence 7/16-21: Vancouver, BC, Canada ISA World Congress of Sociology 7/23-29: Durban, South Africa	21st Annual Meeting of the Society for Conservation Biology 07/1-5: Port Elizabeth, South Africa Public Management of Urban Change in Transitional Cities 07/2 - 13: Budapest, Hungary STREMAH 2007 Tenth International Conference on Studies, Repairs and Maintenance of Heritage Architecture 07/4-6: Prague, Czech Republic ECRM 2007: European Conference on Research Methodology for Business and Management Studies 07/9-10: Lisbon, Portugal In Search of Reconciliation & Peace in Indonesia Workshop 07/19-20: Singapore Symposium on Applied Perception in Graphics and Visualization 07/25-27: Tübingen, Germany	JULY
American Sociological Association 8/5-8: San Francisco, CA, USA Prague Topological Symposium 8/13-19: Prague, Czech Republic International Congress of Mathematicians 8/22-23: Madrid, Spain Diaspora experiences: German-speaking immigrants and their descendants 8/24-27: University of Waterloo, Ontario, Canada	2007 Conference of the International Ass'n for the Study of the Commons 07/31-08/3: Corner Brook, Newfoundland, Canada Infectious Disease: 8th Annual Review 08/2-12: Copenhagen, Denmak American Sociological Association 08/11-14: New York, NY, USA 42nd European Marine Biology Symposium 07/27-31: Kiel, Germany	AUG
Fourth Meeting on Celestial Mechanics - CELMEC IV 9/11-16: San Martino al Cimino, Viterbo, Italy	12th European Congress of Ichthyology (ECI XII) 09/9-13: Dubrovnik, Croatia 7th International Conference on Modelling in Medicine and Biology 09/10-12: The New Forest, UK 15th International Conference on Aquatic Invasive Species 09/23-27: Nijmegen, The Netherlands ICEG 2007: International Conference on e-Government 09/27-28: Montreal, Canada Fourth International Conference on Social Sciences (ICSS 2007) 09/28-30: Barcelona, Spain Oceans 2007 09/29-10/4: Vancouver, British Columbia, Canada	SEP
International Academy of Business and Economics 10/15-18: Las Vegas, NV, USA	International Business & Economics Research Conference 10/1-5: Las Vegas, NV, USA International Academy of Business and Economics 10/14-17: Las Vegas, NV, USA CICKM 2007: International Conference on Intellectual Capital, Knowledge Management and Organisational Learning 10/15-16: Cape Town, South Africa OD Network Conference 2007 ... more here 10/21-24: Baltimore, MD, USA	OCT
Ass'n for Public Policy Analysis & Management 11/2-4, Madison, WI, USA	Association for Public Policy Analysis and Management 11/8-10: Washington, DC, USA 13th Conference of the International Graphonomics Society (IGS) 11/11-14: Melbourne, Australia	NOV
	Climate Impacts on Oceanic Top Predators 12/3-7: La Paz, Mexico Multiphysics 2007 12/12-14: Manchester, UK 2nd International Conference on Mathematics: Trends and Developments 12/27-30: Cairo, Egypt	DEC

Barry Wellman

Ties & Bonds

BBS

Jeremy Daniel Mische Gibson was born 10Dec05 to David Gibson (Soc, U Penna) & Ann Mische (Soc, Rutgers).... Tom Snijders has received a half-time professorship at Nuffield College, Oxford for statistics in the social sciences. He continues living and working in Groningen at other times..... Emmanuel Koku now Asst Prof of Soc, Drexel U.... Paulette Lloyd now Asst Prof of Soc, Indiana U.... A party was held at the National Oceanography Centre in Southampton, England, 21-22Sept06 to honour Peter Killworth who has diagnosed with motor neuron disease (also known as amyotrophic lateral sclerosis/ ALS). Peter, you know you have all of our love and best wishes, but it doesn't hurt to repeat it. And it was great to see you at the Vancouver Sunbelt 5/06.... Eytan Adar and associates (U Washington) have won a Microsoft Live Labs grant for "Vinegar: Leading Indicators in Query Logs" while Lada Adamic & Suresh Bhavani have won one for "VISP: Visualizing Information Search Processes".... David Tindall, Jeffrey Cormier and Mario Diani have received a grant from the Social Science and Humanities Research Council of Canada: "Linking Framing and Social Network Analysis in Social Movements Research." Merrijoy Kelner and Bev Wellman have received a grant from the same agency to hold an international planning conference to study integrative medicine (which links official doctors/hospitals and alternatives such as naturopathy).... Carolyn Mullins passed away April 06. Not only was Carolyn a founding member of the network network at Harvard in the mid-1960s, she participated heavily in the work of the late Nick Mullins in social network theory and studies of scholarly networks. She had a huge impact on my life (and on others) through her workshops and books on how to write clearly in the social sciences.

Is Economics Becoming Networked?

Tom Schelling (MIT) & **Robert Aumann** (Hebrew U) have won 2005 Nobel Prize in Economics for their game theory work. The Royal Swedish Academy noted its usefulness for "security and disarmament policies, price formation on markets, as well as economic and political negotiations." We know it is useful for much more. Schelling points out that "a very small preference not to have too many people unlike in the neighborhood, or even merely a preference for some people like you in the neighborhood ... could lead to such very drastic equilibrium results that looked very much like extreme separation." [Financial Times, 17Dec 05].

Does this mean that economists are being forced to realize that there is action beyond the individual? I dunno. A year ago (10/05), I heard a bunch of Toronto business school graduates being discomforted by Ron Burt (visiting from Chicago) showing how being in brokerage situations is associated with individual success.

Founding Mothers and Fathers

Elizabeth Bott Reminisces: Our founding mother, Elizabeth Bott Spillius, has an article in The Sociological Review 53, 4 (2005): "Anthropology and Psychoanalysis: A Personal Concordance." It's part of a festschrift for Ronald Frankenberg. Here are some excerpts

"I started to become an anthropologist when I was 18, living in Toronto, Canada, when my then boyfriend, Erving Goffman, got me to read Emile Durkheim."[p. 658].

"'Go away and write a novel', said Max Gluckman when I presented my early findings [about networks and family structure in London] at a seminar at [the University of] Manchester." [p. 661].

"Eventually after much painstaking work and sitting hopelessly looking at the data and knowing there should [be] a way of understanding it, an idea floated into my head from nowhere. I had that Archimedes feeling. I remember silently saying ... 'I don't know who you are or how you thought of that, but thank you very much.'... A particular thrill was that an anthropological colleague (Barnes, 1954) had thought of a very similar idea when analysing a very different social situation, a Norwegian fishing village. [p. 662]." [BW: Bott's dissertation and book became Family and Social Network (1957). Details follow in the paper on the ideas of the book, which should be familiar to all readers.]

"[The book] was finally published in 1957, but to be honest I was already changing direction. I was gratified that the book had such a large impact, and that network approaches were taken up both in Britain and abroad. However, even though I did write a long afterword about network methods to the 2nd edition, published in 1971, I only did this so that I could claim copyright on the book, since the Tavistock had copyright on the first edition. I employed a researcher to do much of the ground work for this afterword, and found it really quite painful to write. My interests had shifted … [to] psychoanalysis." [p. 663].

"When I returned from Tonga, … I thought I would be expected to continue working on families, which I did not want to do, and that network research would probably take a new form that I would not enjoy. (I think I was proved right when I read some of the more quantitative studies which began to emerge.)" [p. 663].

"[In this paper,]I have tried to show that although I did not do new anthropological fieldwork after the 1950s, I did not desert anthropologists. Those ideas and excitements have coloured the way I subsequently practiced psychoanalysis." [p.670].

Andre Gunder Frank died April 23, 2005 in his adopted Luxembourg home, after long battles with cancer. Despite illness, Gunder kept working until 2 weeks before he died. Some folks might not consider Gunder to be a network analyst, but I do because of his centrality in the thought and work of world systems folk, including coining the phrase, "the development of underdevelopment". Moreover, Gunder hung out with social network folks in Toronto in the late 1990s-early 2000s, and married one: Nancy Howell. Beverly Wellman and I were the "best people" at the wedding.

Gunder was born in Berlin (1929), his family soon fled the Nazis, and Gunder attended Ann Arbor H.S., Swarthmore Col. and received a PhD in Economics from U Chicago (1957). His career was varied, including an early appointment at Michigan State U, leaving for 10 years in Latin America including being Allende's advisor in Chile (the heyday of fighting against globalizing underdevelopment) which led inevitably to being expelled by the Pinochet regime in 1973. He then had a variety of appointments in Europe, Canada and the U.S.

Gunder published 40 books, and wrote > 1K articles and chapters. When I knew him best in the late 1990s, he was especially proud of his work in non-Eurocentric cycles of development. He was gleeful that his last book, ReOrient: Global Economy in the Asian Age, pointed out the flourishing of Chinese economic dynamism centuries before the 21st century's march of manufacturing from America and Europe to China. (As I write, I am in Los Angeles on leave, where folks tell me that hazardous waste is now the largest export from L.A. harbor to China.) Gunder was always passionate and usually cantankerous, but was also warm and caring. I miss my conversations cum debates with him.

Anatol Rapoport was the subject of a nice story by Jean Drèze in Peace magazine, 10/05: 6. Here's an excerpt: "Back in the 1950s and 1960s, when most game theorists were working for the military establishment and its offshoots, Rapoport (himself not only a game theorist, but also a distinguished psychologist, biologist, philosopher, mathematician, systems theorist and musician) attempted to take the discipline in a completely different direction, oriented towards conflict resolution. His book, Strategy and Conscience, published in 1964, still makes illuminating reading today. Late on, Rapoport played a crucial role in building the foundations of peace science, a unique fusion of science and ethics. In his writings, which have had a deep influence on what follows, one tastes the true joy of scientific enquiry oriented towards human progress – not only material but also ethical." [BW: No wonder Anatol was investigated in the 1950s by homeland security types.]

Alvin Wolfe (U S. Florida) also had a retrospective piece. You can read it in the UrbAnth-L online list, 11Mar06. Here's an excerpt:

"In the early 1960s my studies of the problems of new African states … led me to appreciate the importance of multinational enterprises in the mining and metals industries – not so much in their individual actions as in their systematic organization at a supranational level. My 1962 paper, 'The Rules of Mining in Southern Africa', was the first presentation of the network of corporations that is the 'team' of the title. A 1963 paper, entitled 'The African Mineral Industry: Evolution of a Supranational Level of Integration,' is the first where I recognize the development of a supranational system as a major evolutionary saltation…."

BW: While speaking of southern Africa, remember J. Clyde Mitchell's pioneering "The Kalala Dance" about men from various tribes dancing together on weekends? Those of you who can find the 2005 movie, The Swenkas, directed by Jeppe Ronde, will see Zulu men in Johannesburg who are engaged in a ritualistic fashion show know as the "swanking," as they dress up and compete for prizes (via NY Times review, 10Nov05).

Getting What They Deserved

James Lincoln & **Michael Gerlach** won the Economic Sociology section of the American Sociological Assoc's Viviana Zelizer Distinguished Scholarship Award (2006) for Japan's Network Economy (Cambridge U Press). See a recent Social Networks for Yuki Yasuda's strong review essay based on this book.

Jon Kleinberg (Comp Sci, Cornell U) has won a Macarthur Foundation "genius" award. The ComputerWorld story announcing this emphasizes Jon's contribution to understand web networks and social network structure (24Oct05).

Vincent Lemieux (Pol Sci, Laval U) elected to the Order of Canada. (That's the closest you get to a knighthood up here.) He's written extensively in French on social networks and social capitals.

Sigi Lindenberg (Groningen) is now "Sir Sigmund": he's been anointed a "Knight of the Order of the Dutch Lion". That makes at least 4 network knights: Vincent Lemieux, Sigi, Manuel Castells and Frans Stokman.

Peter Monge (USC Annenberg) has won the 2006 B. Aubrey Fisher Mentorship Award of the Int'l Communication Assoc.

The award recognizes scholars, teachers and advisors who have served as role models and had a major impact on the field of communication by their own accomplishments and those of their students.

Anabel Quan-Haase received a Certificate of Appreciation (2005) for her doctoral dissertation from the Assoc. for Library & Information Science Education (ALISE).

Lynn Smith-Lovin has won the ASA's Social Psych section's Cooley-Mead Award for career achievement.

Charles Tilly was awarded the ASA's Career of Distinguished Scholarship, August 05. As a book reviewer once said, "Tilly writes books faster than I can review them." The ASA's award citation says that his "writings have transformed our understanding of politics, contestation and social change more generally. From his influential early work on urbanization and industrial conflict, to his research on collective action, revolution, and state formation, through his recent emphasis on social relations, identity, and culture…[using] a relational view [and a] secure structural foundation." ASA Footnotes, Nov05: 8]

Brian Uzzi and **Bryan Lancaster** won the ASA's Organization, Occupations and Work section's Best Paper award (2006) for "Embeddedness and Price Formation in the Large Law Firm Market," Amer Soc Review 69: 319-44.

Barry Wellman (U Toronto) won two awards this year. In August 06, he received the Robert and Helen Lynd Career Lifetime Achievement Award from the American Sociological Assoc's Community and Urban Sociology section. The citation for this explicitly mentioned his (ok, my) work in developing and studying a social network conception of community, originally in meatspace and now integrating meatspace and the Internet.

A month later (Sept 2006), Wellman was awarded his Sociology department's only endowed chair: "The S.D. Clark Chair". Who was S.D. Clark? Now deceased, he founded sociology at U Toronto and, in fact, appointed Wellman as Asst Prof way back in 1967.

Chris Winship has won the ASA's Methodology section's Paul Lazarsfeld award, 2006.

Science Networks (which is different from "network science")

Primate Communication may have co-evolved with social bonding. A meta-analysis showed strong relationships between the size of vocal repertoire and both group size and the amount of time spent grooming. [Karen McComb & Stuart Semple, Royal Society Biology Letters, DOI: 10.1098/ 4wbi.2005. 036)

Animal Learning Networks: Two studies have shown that killer whales & chimps pass on to others cultural learning about feeding. Michael Noonan (Canisius Col, NY) found imitation among Marineland killer whales in luring gulls into their swimming tanks. Andrew Whiten, et al. (U of St Andrews, Scotland) found chimps passing on info on how to use sticks to get food even when it was sub-optimal. [NewScientist. com 27Aug05]

The Internet as Jellyfish: Researchers report using graph theory to show that the web "with its central nucleus of nodes, highly interconnected group outside the nucleus and another group of isolated clusters connected directly to the nucleus, resembles a jellyfish. The nucleus "consists of about 15K nodes, and the simple tendrils contain about 5K nodes." [IST Results, 13Oct06].

"***Close or Far: Many Networks Look the Same***" is the surprising (to me) sub-headline of an article by Erica Klarreich in Science News Online (167, 5; 05Jan29). "In recent years, researchers have found that a surprising range of networks has a common structure: a few major hubs with many connections and many minor nodes with only a few connections." "'It's a fundamental advance,' says Albert-Lázló Barabási, a physicist who studies networks at the University of Notre Dam." "Researchers have identified self-similarity in 4 types of complex networks: the World Wide Web, a network of actors who have been in films together, networks of proteins with links between those that can bind to each other, and networks of other cellular molecules with links between molecules involved in the same biochemical reactions." [BW: OTOH, I can show you lots of networks of real people in real situations that do not look the same – or like this.]

Terrorism Networks (if you think I have overloaded the column with this stuff, I left out 3x as many pieces)

Cash in on Terrorism: Neumann College (Aston PA) is offering a certificate program in "Intelligence Analysis" that it says will make you "eligible for intelligence analyst jobs with national, state and local law enforcement agencies". For details: www.neumann. edu. And tell 'em that Osama sent you. [Philadelphia Inquirer, 25Aug05]. If you're reluctant to leave the house, Long Island U (near NYC) is offering an online Master's Program in Homeland Security.

Meanwhile the city of Toronto is getting the secretariat of the Egmont Group, "an organization of 101 of the world's financial intelligence units." [Cdn Dept of Finance, 7July06 press release]. I wonder if the government made any overtures.

Communicating by Unsent Emails: Want to communicate with cell mates without coming to the authorities' notice? Set up an email address / password that is known to all your cell mates. Then type a message into it, but don't send it. Your mates can then login to the email address and read the unsent emails. Reportedly used by the Spanish train bombers. [Intelligence, Number 478, 1May 06]

Want "the largest database ever assembled in the world"? USA Today (11May06) says that the USA's National Security Agency's goal is to create a database of "every call ever made" within the US borders". This follows on news that the NSA is interested in "pattern analysis" on calls within the US and those to and from interesting countries, such as Afghanistan. (New York Times, 24Dec05).

Although I don't play with the spooks, I once talked with a scientist at a US phone company about such data. There were so many terabytes that my brain boggled. I was interested in what exchanges/localities were connected to what others. The project foundered, among other reasons, by the fragmentation of the US phone system. So many calls go to other carriers, and the originating company loses the data at that point. And now there are mobile phones, Internet phones, etc.

Viral Network Immunization of Computers? Eran Shor & associates at Tel-Aviv U have proposed building a network of "honeypot" computers that would attract and analyze viruses, and then quickly transmit the anti-virus solution to others thru a dedicated and secure network. Such computers would then be hubs, distributing the solution to others in their own networks. Simulations suggest that the idea scales positively and non-linearly. A net with 50K computers of which 0.4% are honeypots would see 5% of the net infected before the anti-virus kicks in. By contrast, a 200M computer net with the same % of honeypots would see 0.001% infected. [Nature Physics, DOI: 10.1038/nphys177].

Counterinsurgency: According to a document of this same name, the US Army has recently figured out that the fight in Iraq and Afghanistan is with networks, rather than groups (Version FM3-24; June 06). The 10 page Appendix E is called "Social Network Analysis" which it says "is characterized by a distinct and unique methodology for collecting data, performing statistical analysis, and making visual representations". It says that SNA was especially helpful when used by the US Marines in capturing Saddam Hussein and "the calming of the Fallujah region" (p. E-1). I always thought that it was their leveling of the city that chased folks away. Only 1 paragraph (on E-10) deals with SNA as a perspective rather than a method.

Meanwhile, US Army majors Brian Reed and Scott Efflandt argued in 2001 that "for those leaders at the tip of the spear, an academic grounding in sociology may be the most efficient and useful collegiate specialization." Reed was the primary planner for "Operation Red Dawn", the military operation that apprehended Hussein. "Developing the Warrior-Scholar." Military Review, July: 82-89. [BW: Did the name makers know that in the silly movie Red Dawn, the resistance network overcame the formal armed forces?]

Beware of Greeks Turning on the Taps: Something to think about for the Corfu Sunbelt. Seems that the Greek government mandated that Vodafone make its mobile phone network tappable. And it has been not only for government spooks but for bad guys who hacked in. When confronted with this information, the Greek authorities "denied the possibility that the culprits could be Greek, on the theory that Greek geeks lack the technical knowledge necessary to pull off such a sophisticated hack." [Johna Till Johnson, "A Case of Wiretapping Gone Awry," ComputerWorld Canada, 26May06: 10].

"Right Questions Key to Data Mining" is the headline from the Chicago Tribune, 12May06. Guess SNA is not just a method after all. I look forward to a headline saying: "Right Questions Key to War Starting".

"The Dangers of Social Network Analysis" is the title of the Daily Kos blog, [written by "beerm,"15May06]. Taking off from the US NSA's surveillance, it warns: "Social Network Analysis, despite its academic and impersonal sounding name, is probably the most dangerous use of this information and is a far greater danger to our democracy than the monitoring of individual phone calls". That's because it looks at links, rather than at mere individuals. The blogger warns that the next time NSA actions get revealed, this very agency will look for high network traffic to ID dissidents, or that high-value nodes within the business community will be targeted with newsletters from the ultra right-wing Scaife Foundation. "The misuses of personal information ... has the capability to destroy individuals. The misuse of Social Network Analysis has the capability to destroy individuals and the communities in which they live and work." INSNA and the Wikipedia article on SNA are identified as "a couple of good resources". MiGod, does this mean that the NSA hasn't discovered Wasserman & Faust yet?

MySpace, The NSA's Space: The NSA is funding research into semantic web technologies that "could extract meaning from the mountains of personal data posted on social networking web sites" and combine this with information from banking, retail and property records. The work will benefit from the Semantic Web's common Resource Development Framework which can turn the web into "a kind of universal spreadsheet". [www.newscientist tech. com/article/mg19025556.200 .html].

However, odds are that the variety of language and life will seriously impact surveillance activity. Meanwhile, that old-line blog (aka e-newsletter), Government Computer News warns that even the NSA is hampered by the lack of massive real-time online storage. (www.gcn.com/print25_13/40827-1.html).

Six Degrees Getting Hot

"Six Degrees Medical Consulting" says it is "1 of Canada's leading pharmaceutical communications practitioners, specializing in medical comm., PR and clinical research." And in Oct 06, I met the head of a marketing company also called "Six Degrees" who didn't have a clue who Stan Milgram was, much less Russ Bernard, Peter Killworth or Duncan Watt. "Didn't Marconi come up with the term first?" she asked.

6 Degrees on TV: ABC TV network in the US has a new show (Fall 2006) called, "Six Degrees." Its press release asks: "Who will you touch? Who will touch you? They say that anyone on the planet can be connected to any other person, through a chain of 6 people, which means that no one is a stranger... for long.... This intriguing tale of intertwined destinies...." Can you spot the false premise and grammatical goofs in this release? [The Dec 06 movie Babel also explores the interconnectedness of strangers.]

Lonely Planet Degrees: Also using the concept, the Lonely Planet folks (who do great travel books) brings "Lonely Planet Six Degrees" to the TV world. Their press release says that it "explores the world's coolest cities by connecting with the people who live in them." Each "journey begins with a traveller arriving in a new city with just a single point of contact. From this initial encounter a chain of connectivity is forged across the city as 1 person leads to another and another another." [sic]

Red Auerbach Number: The great Boston Celtic basketball coach died Oct 2006. At least 25 current NBA coaches and 5 general managers have a direct connection to him, And then there are the indirect ties, such as former player, TV analyst and current Celtics coach Doc Rivers who played for Pat Riley who played for Bill Sharman who played for Auerbach.

Abramoff Number: NY Times columnist Paul Krugman introduced a trivia game, "Two Degrees of Jack Abramoff," tracing politicians who have been linked to the convicted influence pedlar. "Grover Norquist, the powerful antitax lobbyist, is a 1-degree man" because he was Abramoff's campaign manager when he ran for chair of the College Republication National Committee. Karl Rove, the president's political advisor, is a 2-degree man because he hired Abramoff's assistant as his own assistant, as is former Republican House Majority leader Tom DeLay. [26Sept05: "Find the Brownie"]. These are all Republicans; I remember in the 1950s when those on the left refused to believe in "guilt by association".

MySpace as a Virus: Reportedly teenager "Sami" wrote an AJAX worm and put it on his MySpace profile. It caused anyone who looked at his site to "friend" him and propagate the worm on their own pages. Within a day, Sami had > 1M new "friends" [Quinn Norton, "Beguiling but Beware." Wired News, 3Oct06].

Negative Networks: "In the office in which I work, there are 5 people of whom I am afraid. Each of these 5 people is afraid of 4 people (excluding overlaps), for a total of 20, and each of these 20 people is afraid of 6 people, making a total of 120 people who are feared by at least 1 person." [From Joseph Heller's novel, Something Happened, as quoted in Report on Business Magazine, Toronto, May 06: 78].

Six Apart makes really good blogging tools (Moveable Type, Typepad) plus running an adult-oriented blog, Vox, and a well-liked teen-oriented one, LiveJournal. Although my friends have used the tools for years, it took The Economist's 25Nov06 story to make me aware that the sum is greater than the parts. For one thing, Vox is the only blog I know of that allows bloggers to specify who gets to read what.

Networked Publications

Management and Organizational Review is the name of a year-old journal dedicated to publishing China-related studies, both theoretical and empirical. The issues I've seen are high quality. Networkers Yanjie Bian (HK U of Science & Technology) & Joseph Galaskiewicz (Soc, U Arizona) are senior editors. Info: www.iacmr.org/MOR.htm

Structure and Dynamics was announced Sept 05 as an e-journal for anthropology and related sciences, especially cross-disciplinary research. Networker Douglas White (Anthro, U Cal Irvine) is the editor in chief.

Networks and Heterogeneous Media was announced Jan 06 as a new applied math journal. Social networks is included in its list of topics. Info at: http://cpde.iac.cnr.it/Convegno_NHM/aim.php.].

Short Schticks

Network Survey Cache: David Tindall & Todd Malinick are developing a web repository at the Anthro/Soc Dept, U British Columbia of survey instruments for collecting network data and associated publications. Contact them: tindall@interchange.-ubc.ca

The Weakness of Board Ties: Antonio Villar, who used to be the largest benefactor of the (NY) Metropolitan Opera, says that once he was under fraud indictment, no other Met board member ever contacted him to offer help or even to express sympathy.

Networking Communication Research was the theme of the International Communication Assoc conf in Dresden May/06. Not surprisingly, Ronald Rice (U of California - Santa Barbara) was ICA's President-elect and conference chair.

God as an Agent-Based Networker: "God in his infinite freedom continuously creates a world that reflects that freedom at all levels of the evolutionary processes to greater and greater complexity. He is not continually intervening, but rather allows, participates, loves." Rev. George Coyne, director of the Vatican Observatory, as quoted by Nicole Winfield, "Vatican Official Refutes Intelligent Design" [AP, 18Nov05].

SNA as a Hot Area: In mid-Aug 2006, Gartner.com identified social network analysis as one of 4 areas it thinks will have the great impact on businesses over the next decade, forecasting it will reach maturity in 2 years. Gartner sees SNA as using the information and knowledge gathered from people's personal networks to identify target markets, create project teams and discover unvoiced conclusions. Gartner says SNA involves "collecting massive amounts of data from multiple sources, analyzing the data to identify relationships and mining it for new information." [Antone Gonsalves, Information Week, 9Aug06].

Cleaning up Wikipedia: Which reminds me: While I've been impressed by almost all of the Wikipedia entries I've seen recently, the one on "social networks" is badly contaminated by proponents (vendors?) of social networking software (such as MySpace) touting their virtues. Would someone please clean this entry up, and move the social-software folks to their own sandbox.

Sing a Song of Networks: There's a song CD out called, "The Strength of Weak Ties" by the group called Lotus. Lotus is an instrumental jamband that "splices light electronica sounds and standard-issue Phisheadry, samples and strums; jazz and funk-heavy world beats.... [This $16 13disc] is too much of a combination platter to truly hold a new listener's interest ... as the songs drift and dip and float from genre to genre". [David Berger, "Lotus: The Strength of Weak Ties" Harmonium Archive, 05May06]. Perhaps this is why Mark is not claiming royalties or trademark violation.

The Center for Collective Intelligence is a new MIT outfit that hopes to use "large numbers of people" to solve business, scientific, and societal problems. Director Tom Malone says that its basic research question will be: "How can people and computers be connected so that, collectively, they act more intelligently than any individuals, groups, or computers have ever done before? One of the center's first projects will be a Wikipedia-style business book about the effects of social networks on business operations." [Chronicle of Higher Education, 13Oct06. See also http://cci.mit.edu].

Ancestral Networks: Steve Olson, et al. calculate that every person who was alive 5K-7K years ago was an ancestor to all 6B people living now, or their line died out earlier. [In Mapping Human History, as described by Matt Crenson, "Roots of Human Family Tree are Shallow," AP, 1July06].

Always Talk to Strangers is David Wygant's guide to single (American) adults wanting to find a partner (Perigee, 2005). It includes tips on a mental and psychological makeover. [BW: What about sociological?] . Perhaps it is coincidence, but Wharton Business School's "Strategic Management" newsletter had a 2005 article, "Do Talk to Strangers: Encouraging Performative Ties to Create Competitive Advantage" – through "impromptu communications made by colleagues who are strangers in which critical knowledge is transferred with no expectation of a quid pro quo. The advocate is Sheen Levine, from Singapore Management University.

Networking for Fun and Profit: When entrepreneur Donna Messer (head of "ConnectUS Communications Canada" works a room, she uses these tips: 1. Carry plenty of business cards. Wear a jacket with 2 pockets – right pocket to carry your cards, the left pocket to collect others. 2. Quickly scan name tags while looking around but not while talking to someone – always maintain eye contact. 3. Look for people on their own, so that you can interact one-on-one. (Just like in my high school dances). 4. Ask for a business card before offering your own – it's less presumptuous. 5. Try to enlist a mutual acquaintance to make an intro to someone you'd like to know – but don't tell Ron Burt.

Not a Retiring Sort: I asked Elihu Katz (8/05) when he was going to retire, he answered: "I am too busy working to think about it!" When he gave permission to quote his non-retirement line, he asked that I also present his bon mot at the American Political Science Assoc. conference, 8/05: "Bush can't blame God," to which Kathleen Jamieson added, "But God can blame Bush!". He also has a new intro to the legendary Katz & Lazarsfeld, Personal Influence, which discusses social networks.

CONNECTIONS 27(2): 15-23
http://www.insna.org/Connections-Web/Volume27-2/Lin.pdf

The Blog Network in America: Blogs as Indicators of Relationships among US Cities

Jia Lin[1]
Department of Communication; State University of New York at Buffalo

Alexander Halavais[2]
Department of Communication; State University of New York at Buffalo

Bin Zhang[3]
School of Medicine; University of California Los Angeles

An analysis of links among U.S. weblogs is done to examine the interpersonal social network and social connections among U.S. cities. Drawing 4,241 weblogs from the NITLE census dataset that are identified as being located in the United States, this project extracts the outward links of these weblogs and uses them to analyze the relationship between cities. A total of 632 U.S. city/region units, represented by the first three-digits of US postal codes, are taken as nodes of the network. In total, 41,212 permanent links from blogs of each of the city units are counted as weighted arcs in the network. Inlinks and outlinks of each city unit are recorded for analysis. The study finds that the city units whose bloggers attract most inlinks are Manhattan, San Francisco and Bay Area, Washington, D.C. and its western suburbs, Boston and its suburbs, Los Angeles and Seattle. The study discovers a super-metropolitan cluster, transcending geographical boundaries, within which the cities traditionally associated with cultural elites are closely connected. For other less metropolitan areas, blogs are most heavily connected at a geographically local level, and then extend to a national network.

INTRODUCTION

Weblogs, or *blogs*, are self-published websites that have burgeoned since the late 1990s and by December of 2004, the number of blogs had grown to 7 million (Technorati.com). Spontaneous, self-reported expressions made conveniently available online priovede great opportunities for social-science research. The blogosphere, the totality of interconnected blogs, provides two layers of information: content and relationships. Writings in weblog entries archive people's everyday experience, while hyperlinks among individual blogs trace some form of social structure. Bloggers are not only noting down their experiences and thoughts, but also trying to reach out to broader audiences, share opinions and to manage their personal knowledge base.

The digital revolution has profoundly redefined the dynamics between space and place. Though people may remain physically stationary, their identity, social capital, and flows of communication often exist in a spatial form. On the other hand, what people bring to online communications is inevitably shaped by "their gender, stage in life-cycle, cultural milieu, socioeconomic status, and offline connections with others" (Wellman & Gulia, 1999). While blogging is an on-line activity that transcends geographical boundaries, the self-images presented to the public and the hyperlinks used are shaped by who the bloggers are in real life, including their physical location.

This project explores social connections among American cities

[1] jialin@gmail.com

[2] alex@halavais.net

[3] binzhang.ucla@gmail.com

by examining weblog hyperlinks among them. We postulate that the range and density of networks of social relations reflect the strength of social connections among cities. By studying the network of cities linked through hyperlinks on blogs, this project will identify the most influential cities in the country, as well as the geographical distribution of city clusters connected by blog links.

LITERATURE REVIEW

Weblog networks as indicators of social connection

The blogosphere provides two layers of information: content and relationships. Compared to personal journals, blogs engage a public readership through posting online and linking to other sites. According to Rebecca Blood (2000), blogs started as link-driven sites and each was a mixture of links to news, commentary and thoughts. Blogging makes use of a set of practices and protocols (pinging, subscribing, commenting, trackbacks, and hyperlinking) through which bloggers co-construct a set of affiliation networks, within which individuals with the same interests or close social ties interact with and refer to one another. MP

Tens of millions of personal weblogs have become an indispensable part of World Wide Web. On-line hosting services like Blogger, TypePad, and Livejournal, have made blogging easier for internet users and have stimulated a wider adoption of blogging. According to National Institute for Technology & Liberal Education (NITLE) census data, from July to November 2003 the blogs on *blog-hosting sites* such as *Blogspot*, *Livejournal* and *Diaryland* increased more than 4 times from 176 thousand to 720 thousand. It is estimated that there are at least as many *stand-alone blogs*, blogs that are self-hosted or at least have their own domain names. In either case, bloggers continue to update and post new entries regularly, leading to a large and quickly-growing network of hyperlinked sites that form what has been termed the *blogosphere*.

Broadly, a weblog is any "hierarchy of text, images, media objects and data, arranged chronologically, that can be viewed in an HTML browser" (Winer, 2003). This loose definition can be applied to almost any sort of regularly updated website, but what makes weblogs special is that they are the "unedited voice of a person." Weblogs can be categorized into one or more of three types: first, a personal diary recording daily activity and thoughts; second, a collection of links to other websites worthy of recommendations, usually with a few words of comments; and third, a forum devoted to specific topics. Beyond personal blogs, there are also a significant number of community weblogs that are coauthored by a number of contributors; this includes sites like Slashdot and Boing Boing, among others. Some corporations and other organizations—from Microsoft to Ford to Boeing—also maintain weblogs.

The depth and size of the content, as well as its relatively easy accessibility, make weblogs a potentially valuable resource for social-psychological studies. Blogging technology breaks down traditional centralized authorship, enabling everyone with internet access to become a potential author. The blogosphere provides a reflection of ongoing distributed discourse (Halavais, 2002). Political blogs — among the most popular — have been the subject of several studies related to political agendas and public opinions in the US (Lin & Halavais, 2005; Adamic & Glance, 2005; Cornfield et al., 2005). In the business world, personal blogs have become an useful source for marketers and advertisers to study word of mouth communication and detect new trends and consumer behaviors.

The blogosphere provides two layers of information: content and relationships. Compared to personal journals, blogs engage a public readership through posting online and linking to other sites. According to Rebecca Blood (2000), blogs started as link-driven sites and each was a mixture of links to news, commentary and thoughts. Blogging makes use of a set of practices and protocols (pinging, subscribing, commenting, trackbacks, and hyperlinking) through which bloggers co-construct a set of affiliation networks, within which individuals with the same interests or close social ties interact with and refer to one another.

For the purpose of regular reading, many bloggers place links of other blogs on their index pages (permanent links, also sometimes called "blogrolls") or subscribe to various blogs through syndicated feeds. In this sense, one cluster of hyperlinks could be viewed a virtual neighborhood or community (see Kumar et al, 1999). Such groups may be driven by political preference, culture or academic interests, health or spiritual support, or by familial and friendship ties. According to Wellman (2001), networks online or offline represent social network of relationships that provide sociability, information and sense of belonging. Hyperlink networks reflect the structure of social relationships online, and can be used in the research of international communication, interpersonal communication and ecommerce (Park, 2003). Blog networks, as an increasingly important component of Internet, can be viewed as another effective indicator of virtual community on line. Because the hyperlinks on blogs are frequently created by individual authors, they represent a more fine-grained view of social structure. The collective linking behavior among geographical units can reflect the overall social connections: do bloggers link to each other regardless of geographical boundaries? Are there stronger bonds among certain cities than others?

Geography in the digital age

The prevalence of telecommunication technologies has generated some popular notions of the fading impact of geographical location and physical distance. From the early prophecy of the "global village" by Marshal McLuhan (1968), to the expansion of the "global city" (King, 1990), from the claim of the "death of distance" by Cairncross (1997), to the design of the "City of Bits" and the "E-topia" by Mitchell (1999), scholars suggest that in the digital age, where people work, live and socialize is not limited by where they physically are anymore. The rise of "world cities" has been widely accepted in urban studies. Cities like New York, Los Angeles, London or Paris have virtually exceeded the national boundaries and become the world's "headquarters" because of their prominent place in corporate control, communication networks, and cultural production (Logan & Molotch, 1987; King, 1991). King (1991) introduced the concept of the world urban system, an interdependent system, consisting of people, knowledge, images and ideas. Such "world cities", according to King, bring together the highly-paid international elites and a transnational producer service class to form a spatial center for global transmission of news, information and culture. In all, these scholars believe information technology enables a time-space compression and the decrease of unidimensional spatial patterns.

Meanwhile, the notion that "geography is dead" is criticized as one of the greatest of the new economy myths. Opposing voices arise to reiterate the importance of geography in a new urban landscape, and suggest that information technology only reinforces community ties. Wellman and colleagues' research (2001) on Canadians' use of the Internet finds that people use email mainly to enhance communications with acquaintance such as kin and neighbors, and communication is lower with distant than nearby friends. Kotkin (2001), in his book on the geographical distribution of the new information industry in the United States, declares that the appeal of a place is an increasing priority when people choose where to work and live, since the communication can be achieved over a long distance anyway. Richard Florida's (2002) study of the creative economy finds that human capital today is more selective when deciding worksites. In the international setting, Halavais (2000) finds that the distribution of the hyperlinks is far from matching the worldwide distribution of websites, and there are clear national borders on the Internet. Barnett and Choi's (1995, 1999) examination of the international telephone network suggests that international telecommunication network is to a large degree determined by the factor of physical location and the network can be clearly differentiated into three subgroups (Latin American, Europe and Asia sub groups), with United States acting as a "liaison" in the center. Zook's (2001) track of internet domain registrations concludes that "the Internet is a more selective network that parallels physical geography and economic development" (p. 3). The growth of the Internet is concentrated in the big cities and urban areas; rather than destroying geography, the internet is selectively connecting a small group of people into highly interactive networks.

The United States is one of the most connected countries in the world with the largest internet population. According to the Pew Internet and American Life Project, 2% to 7% of adult Internet users in the United States are keeping weblogs, of which 10% update them daily. Blogs have become an alternative, grass-roots form of media. Especially when it comes to political news, the blogosphere has been viewed as an indispensable source by mainstream media and internet users. The occurrence of decentralized publishing not only redefines community space, but also redefines the role of authorship and readership. The blog attracting many in-links can be viewed as a credible news source, a popular opinion leader in a certain field or a platform for good writing, while out-links from a blog indicate the seeking of such news, opinion or writings. By relating blogging to the question of physical geography versus cyberspace, this project will examine where the popular authorship located; that is, the degree to which the geographical location of the blogger is related to his or her prominence within the blogosphere. At the same time, we may ask, do such linking behaviors create an integrated network or a fragmented one with subgroup clusters?

Our society is not a collection of random units, but an interactive and hierarchical network consisting of specific geographical relationships that help to define cultural relationships within and between cities. A city is "a state of mind, a body of customs and traditions and of the organized attitudes and sentiments that inhere in these customs and are transmitted with this tradition" (Park, 1984, p. 1). Accordingly, we assumed that people of same place will demonstrate more or less similar patterns in selecting and attaching to social networks. Social network analysis is widely used in examining the structure of a social entity where geographical location is often seen as a node in the network (Brunn and Dudge, 2001; Zook, 2001; Barnett & Park, 2005; Barnett, 1999; Barnett & Choi, 1995; Halavais, 2000). This study postulates that the range and density of social relation networks reflect the strength of social connections among city units.

In sum, this project seeks to answer the following two research questions:

1. To what degree are the hypertextual expressions of blog authors related to the geographical locations from which they blog?
2. Where are the centers of opinion leaders? Are there clusters of opinion congregation, and if so, how are they identified?

METHODOLOGY

Data

Blog samples for this study are retrieved from NITLE census data from June and July of 2003. The database consists of index pages from about 120,000 blogs. Permanent links are retrieved from the blogs that are identified as being authored i8n the US. In most cases, the permanent links appear on the index pages; but for many blogs hosted by *Livejournal* and *Diaryland*, such links appear instead on member information pages. To exclude hyperlinks appearing in periodic postings, which are less likely to be to weblogs and are therefore less indicative of interpersonal affiliation, only URLs placed on sidebars and grouped together are extracted. The links to such popular web services as hosting sites, providers of visitor statistics, commenting services, technology assistance, together with popular mainstream news sites are excluded from the data.

The geographical locations of blogs and the targets of their outlinks are extracted. Computer-assisted automatic retrieval is realized by a custom set of software tools. A crawler searches all relevant web pages, retrieving the source code (including html code and plain text) of every page, and extracts keywords with geographical information. The crawler searched following patterns that contain geographical information:

Geotags: When present, explicit meta-tags pinpointing the geographical location of a site were the most unambiguous indicator of a given site's location. After extracting the values of longitude and latitude from meta-tags, they were mapped to zip code, if located in the United States. Unfortunately, only very small proportion of weblogs provide such meta-data.

Local weather: City location can be inferred by weather– related links, since the more exact location the blogger provides to the weather service, the more precise weather forecast they can receive. *Weather.com* and *weatherpixie. com* are the two dominant weather services used by bloggers. In each case, there is an indicator of the geographic location in the text of the URL; namely, either a zip code or local airport code.

Blogchalk profile: Blogchalk represented personal information about the blogger in machine- and-human- readable forms. Included among these keywords were the home city and country of the blogger. While now largely defunct, the service was still occasionally used among the weblogs surveyed during this period.

Blogger profiles at hosted weblogs: *Blogger*, *Livejournal* and *Diaryland*, the three major blogging host services during the period, provided web pages for user profiles where users were able to list their location.

Using the technologies described above, a previous study on mapping the distribution of weblogs in America generate the map in Figure 1 (Lin & Halavais, 2005).

This analysis made use of data from 2003, which, while slightly dated, provides a good snapshot of the web. The approach taken here may not be as appropriate when applied to the current or future blogosphere, because of the increased use of syndicated feeds (RSS) and aggregators, which reduce the apparent links between blogs.

A total of 4,241 weblogs and 41,212 permanent links are identified for their location in the US. Blogs and links are plotted into their corresponding three-digit zip codes. A total of 632 U.S. city/region units have at least one blog present. The geographical location of blogs has been shown to be consistent with the population distribution and concentrations of high socio-economic status (Lin& Halavais, 2005). Park (1984) notes residential homogeneity as an important indicator in sorting neighborhoods within a city space. Compared to a partition of population in greater metropolitan area which is too general, or into 5-digit zip code corresponding to streets or blocks (Weiss, 1989), which overstates the population variance, 3-digit zip code units represent a middle approach that defines a geographical unit in a way that is widely used in marketing and political targeting strategies.

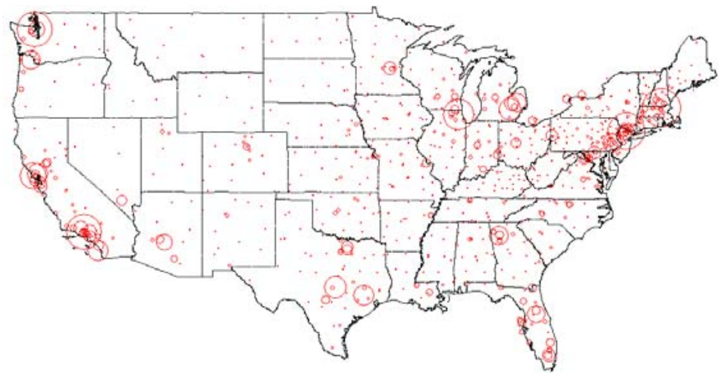


Figure 1. Distribution of bloggers in America. Circle radius is proportional to the number of webloggers in a region.

Data analysis

After collecting all weblogs with links and geographical locations, each weblog i and its n_i out-linked weblogs are transformed into a list of 3-digit zipcodes, represented by

$$l_i = < z_0^i,\ z_1^i,\ z_2^i, ...,\ z_{n_i}^i > \qquad (1)$$

where z_0^i is the 3-digit zipcode for the weblog i and the remaining elements are the 3-digit zipcodes of the weblogs referred by the weblog i.

Given N weblogs, there will be N lists encoded as the formula (1). From these lists of three digit zipcodes, we can construct a linkage matrix $A=[a_{ij}]$, where each entry a_{ij} represents the number of outgoing links from a region represented by a 3-digit zipcode i to a region encoded by a 3-digit zipcode j. The linkage matrix A represents a directed network. Several important network-based measures are the number of incoming links k_{in}, the number of outgoing links k_{out}, and the number of local links k_{local} (the number links to the blogs of the same zip code) and they are formally expressed as follows:

$$k_{in}^i = \sum_{j,j \neq i} a_{ji}, \qquad (2)$$

$$k_{out}^i = \sum_{j,j \neq i} a_{ji}, \qquad (3)$$

$$k_{local}^i = \sum_{j,j \neq i} a_{ji}, \qquad (4)$$

where i and j represent different regions encoded by 3-digit zipcodes.

Notice that the local links are excluded from k_{in} and k_{out} . k_{in} , the total number of inlinks from outside zip codes will be taken as the prestige score.

In order to later perform a cluster analysis, we derive an undirected adjacency matrix $A = [a_{ij}]$ from the linkage matrix A:

$$\begin{cases} u_{ij} = a_{ij} + a_{ji}, & i \neq j \\ u_{ii} = 0 \end{cases} \quad (5)$$

The adjacency matrix U is symmetric and each entry represents the strength of accumulative interactions (total links) between the two corresponding regions. The connectivity of a region to adjacent nodes is indicated by the total number of inlinks and outlinks (excluding the links within the region), given by

$$k^i = \sum_j u_{ij} \quad (5)$$

Additionally, we cluster regions within the network to discover linking relationships. Clusters are subsets of nodes that are tightly linked with each other. For this purpose, we perform an *agglomerative* hierarchical clustering using an average distance clustering algorithm, in which the distance between two clusters is defined as the average number of links between all pairs of nodes in each pair of clusters (Duda et.al, 2000). The hierarchical clustering initially assigns each node to a cluster and then iteratively merges the closest pair of clusters until all nodes are clustered into a single cluster.

The network analyses presented above were implemented in R (Ihaka & Gentleman, 1996). To better visualize the weblog network, we also use the Fruchterman Rheingold algorithm implemented in Pajek to represent the network in 3-dimensional space.

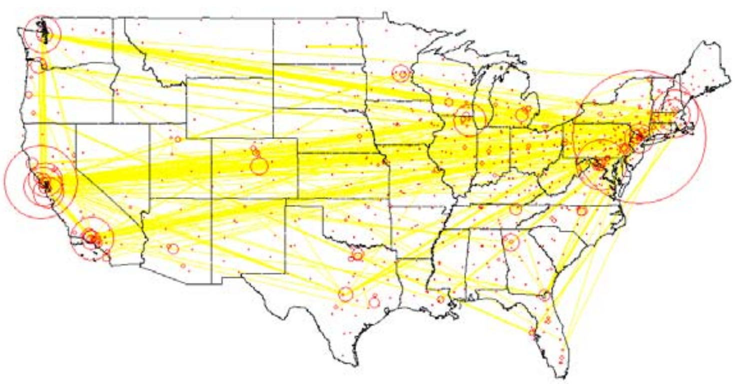


Figure 2. Blog network plot on American map. Circle radius is proportional to the number of inbound links to a region.

FINDINGS

Geographical map of blog links

Blog links are plotted into the American map shown in Figure 2. There are clearly heavy links is across the continent from coast to coast. The bigger circles in areas indicate larger numbers of in-links from other city units.

The ranking of city units by the number of in-links are shown in Table 1. The city units whose bloggers attracted most inlinks include Manhattan, San Francisco and the San Francisco Bay Area, Washington, D.C. and its western suburbs, Boston and its suburbs, Los Angeles and Seattle.

Table 1. Prestige score-City units with the most inlinks (excluding links from local blogs)

Prestige ranks	3-digit zip code	City unit	No. of inlinks
1	100	Manhattan	2657
2	941	San Francisco	1446
3	200	D.C	1017
4	21	Boston	942
5	201	Northern Virginia	885
6	900	Los Angeles	845
7	940	South Bay of SF	778
8	981	Seattle	728
9	945	West Bay of SF	662
10	606	Chicago	634
11	24	Suburb of Boston	634
12	950	Santa Cruz	628
13	112	Brooklyn	573
14	902	West Los Angeles	361
15	554	MINNEAPOLIS	347
16	809	Colorado Springs	345
17	303	Atlanta	342
18	980	West Seattle	322
19	972	Portland	301
20	787	Austin	288

To examine if the number of inlinks are determined by the number of bloggers, we weight inlinks by total number of blogs in each city, resulting in Table 2. Compared to the indegree ranking (Table 1), 7 out of the top ten cities remain the same. This indicates that the blog ties are only partially determined by population size.

Table 2. Ranking of cities with the highest degree of normalized inlinks.

Prestige ranks	3-digit zip code	City unit	Normalized inlinks
1	200	D.C	16.67213
2	940	South San Francisco	15.56
3	100	NewYork City	13.08867
4	941	San Francisco	11.568
5	24	Suburb of Boston	11.32143
6	201	Northern Virginia	10.79268
7	950	Santa Cruz	9.515152
8	112	Brooklyn	9.095238
9	900	Los Angeles	8.894737
10	554	Minneapolis	8.463415

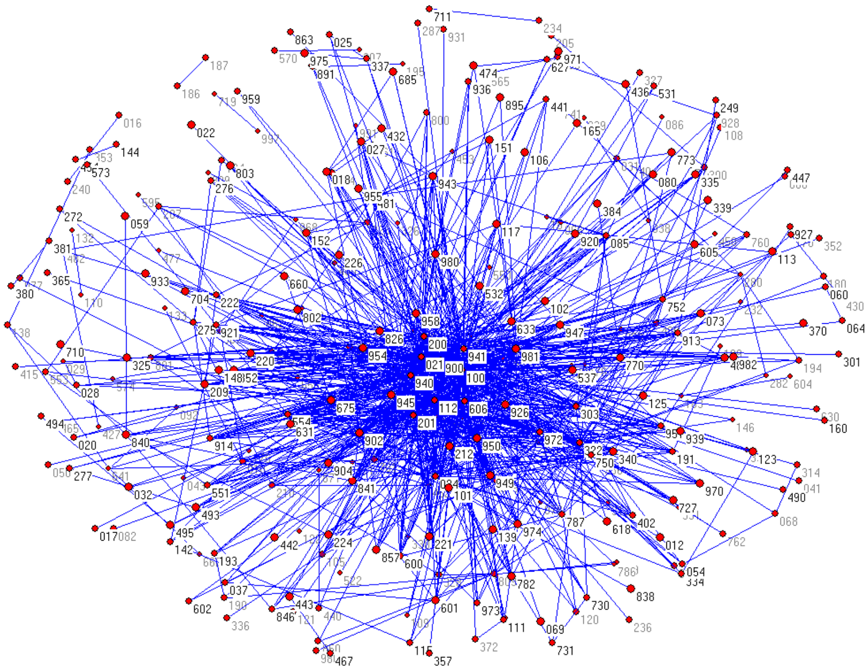


Figure 3. Blog network based on centralities

Figure 3 shows the hyperlink network based on the centrality of each city node. Network density is a measure of the relative number of connection, and it is ranged from 0 to 1. A fully connected network has a density of 1. The density for blog hyperlink network is 0.0662, which indicates that the network is far from highly interconnected.

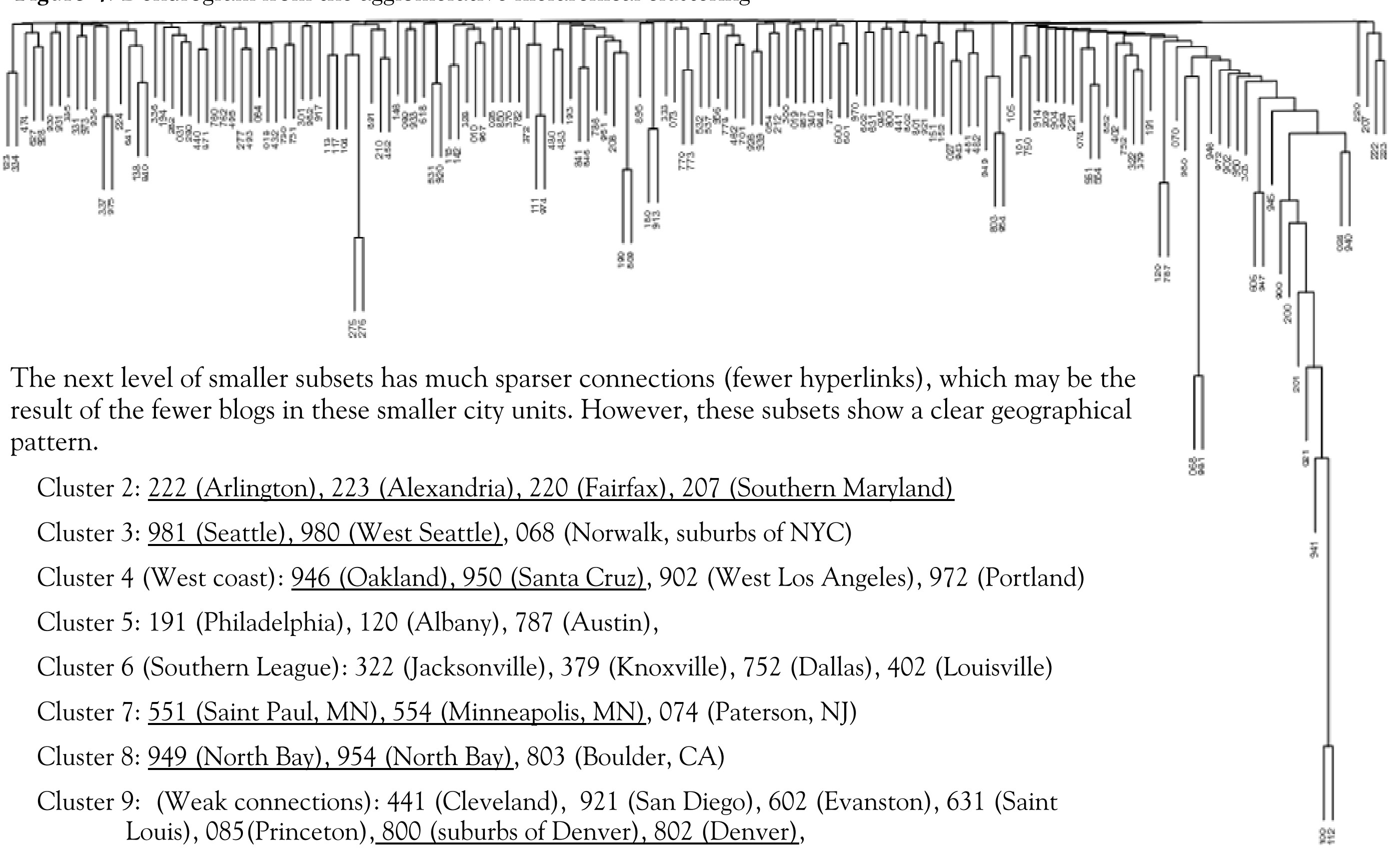
Cluster analysis

The dendrogram from the agglomerative hierarchical clustering is shown in figure 4. There is a large cluster which consists of the major cities in the United States.

Cluster analysis finds about 30 significant clusters and dyads. The largest cluster sits in the center of the network and connects to almost all nodes and subgroup networks (Figure 3). Shown below are the sub-networks grouped by their connections, and the order of the listing is based on the clusters' distance to the central cluster (cluster 1 shown in Figure 5).

Clusters 1(Strongest connection):
101(Manhattan), 112 (Brooklyn), 941 (San Francisco), 940 (south bay of San Francisco), 945 (East San Francisco Bay), 947 (Berkeley), 021 (Boston), 024 (Suburbs of Boston), 201 (DC area), 200 (DC area), 606(Chicago), 900(Los Angeles)

Figure 4. Dendrogram from the agglomerative hierarchical clustering



The next level of smaller subsets has much sparser connections (fewer hyperlinks), which may be the result of the fewer blogs in these smaller city units. However, these subsets show a clear geographical pattern.

Cluster 2: 222 (Arlington), 223 (Alexandria), 220 (Fairfax), 207 (Southern Maryland)

Cluster 3: 981 (Seattle), 980 (West Seattle), 068 (Norwalk, suburbs of NYC)

Cluster 4 (West coast): 946 (Oakland), 950 (Santa Cruz), 902 (West Los Angeles), 972 (Portland)

Cluster 5: 191 (Philadelphia), 120 (Albany), 787 (Austin),

Cluster 6 (Southern League): 322 (Jacksonville), 379 (Knoxville), 752 (Dallas), 402 (Louisville)

Cluster 7: 551 (Saint Paul, MN), 554 (Minneapolis, MN), 074 (Paterson, NJ)

Cluster 8: 949 (North Bay), 954 (North Bay), 803 (Boulder, CA)

Cluster 9: (Weak connections): 441 (Cleveland), 921 (San Diego), 602 (Evanston), 631 (Saint Louis), 085(Princeton), 800 (suburbs of Denver), 802 (Denver),

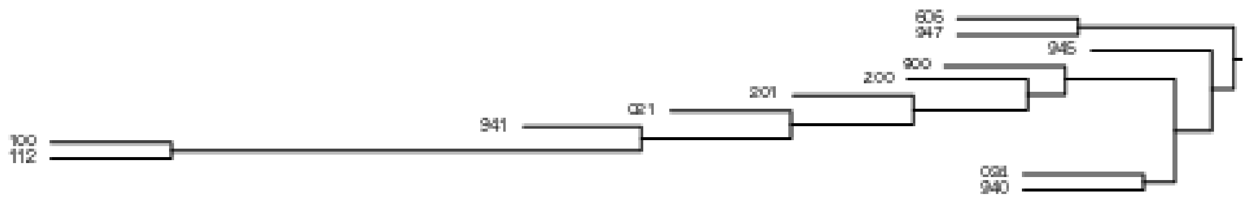


Figure 5. Clusters' distance to the central cluster

Cluster 6 (Southern League): 322 (Jacksonville), 379 (Knoxville), 752 (Dallas), 402 (Louisville)

Cluster 7: 551 (Saint Paul, MN), 554 (Minneapolis, MN), 074 (Paterson, NJ)

Cluster 8: 949 (North Bay), 954 (North Bay), 803 (Boulder, CA)

Cluster 9: (Weak connections): 441 (Cleveland), 921 (San Diego), 602 (Evanston), 631 (Saint Louis), 085(Princeton), 800 (suburbs of Denver), 802 (Denver),

Cluster 10: 926 (Santa Anna), 939 (Salinas), 054 (Vermont), 212 (Baltimore)

Cluster 11(Weak): 973 (Salem), 975 (Medford), 936 (Fresno), 337 (Miami), 331 (St. Petersburg)

Cluster 12: 600 (suburbs of Chicago), 601 (suburbs of Chicago), 727(suburbs of Oklahoma City)

Cluster 13: 770 (Houston), 773 (North Houston), 073 (Newark, NJ)

The following subsets consist of dyadic notes that group with each other exclusively:

Dyad 1: 481 (Ann Arbor), 482 (Detroit)

Dyad 2: 532 (Milwaukee) , 537 (Madison)

Dyad 3: 151 (Pittsburgh) , 152 (Suburbs of Pittsburgh)

Dyad 4: 841 (Salt Lake City), 846 (Suburbs of Salt Lake city)

Dyad 5: 480 (North suburbs of Detroit), 483 (North suburbs of Detroit)

Dyad 6: 275 (Suburbs of Raleigh), 276 (Raleigh)

Dyad 7: 113 (Queens), 117 (Mid Island)

Dyad 8: 730 (Suburbs of Okalahoma City), 731(Okalahoma City)

Dyad 9: 760 (Fort Worth), 762 (Forth Worth)

Dyad 10: 930(Ventura), 931 (Santa Barbara)

Dyad 11: 631 (San Louis), 606 (Chicago)

Subsets with underlines indicate the close geographical distance between the nodes. The list indicates that most of the clusters with strong connections also share geographical proximity.

DISCUSSION

Blogs represent the collection of experiences and opinions of individual internet users. The networks presented in this research are based on dynamics of on-line communications among individuals. Therefore, each node in the network represents the geographical location of individual content providers instead of hosting servers. We assume that people in the same place, measured by 3-digit zip code, tend to possess some more or less collective traits. While individual difference is important, we take the view that the sum of individuals in one city reflects the general trend at the macro level.

This research finds that networks among American cities, as indicated by weblog hyperlinks, are densest among metropolitan cities on the West and the East coasts. Cities with cultural-political prominence, like Boston, San Francisco, New York, Washington and Los Angeles, traditionally seedbeds of national opinion, forge a highly connected cluster in the center of the national networks (figure 4). Meanwhile, satellite cities or suburbs around some of these cities also play a significant role in this central cluster, consistent with migration patterns in recent years. The San Francisco Bay Area has become an active cultural and technological hub, especially since the high-tech boom in 1990s. With the sharp increase of population and rising cost of housing in the area, more middle class and young people, including large number of creative workers, have moved to the suburbs, energizing and urbanizing these traditionally more conservative areas. Blogs in these cities, including surrounding areas, reciprocate hyperlinks and maintain coherent clusters. We call this group of cities the "super metropolitan cluster." In a sense, this cluster transcends geography, though the urban/suburban connection often remains intact. Our research finds that the bloggers in these cities tend to receive more inlinks than those in other areas, beyond what would be expected given the concentration of population. The strong connection among these cities also supports Fischer's observation that for cosmopolitan residents, close friendship are often long-distance, since urban residents move around more frequently during their lifetime and they accumulate more weak social ties (1982).

For the cities of less cultural-political significance, the connections are first and foremost with places that near them. Other than these two types of clusters, we also found less prominent clusters of connections that are based on similar city profiles. For example, the cluster made of Philadelphia, Albany and Austin, and the cluster made of Pittsburgh, Evanston, Saint

Louis, Princeton, Cleveland, Denver and San Diego seem to correspond to the prominent educational institutions and new cultural dynamics in these areas. Regional ties seem also play role in shaping the clusters of cities. The cluster made of Portland, Oakland, Santa Cruz and West Los Angeles, and the cluster made of Jacksonville, Knoxville, Louisville and Dallas seem to reflect a west-coast and a southern block sensibility, respectively.

If blogging makes possible the decentralization of publication and news resources, it might be also prompt the decentralization of large cities. However, when it comes to readership (as reflected in hyperlinked commentary on other blogs), choices made by audiences are not particularly geographically decentralized. The overwhelming number of links pouring into cities like New York, Boston, San Francisco, Washington, and Los Angeles shows that these centers of cultural and news production still attract the most attention nationwide.

This research project is limited by the technical difficulties of extracting valid link data from blogs, as well as their geographical locations. There is a lack of blog and link data from small cities, especially those in the midwest. The scale-free distribution of the network we observe based on these data suggests that a larger data set would yield similar results. Certainly a larger data set may provide more finely-grained information about local connections and structural relationships—which areas are hubs, bridges or tree nodes in the network. Nonetheless, the work here demonstrates the possibility of using links among weblogs to measure socio-geographical relationships, and suggests some interesting national patterns of discursive clustering. Future work that allows for ongoing monitoring of such relationships should provide an interesting barometer of social exchange.

REFERENCES

Adamic L & Glance N. 2005. The political blogosphere and the 2004 U.S. election: Divided they blog. *LinkKDD-2005*, Chicago, IL, Aug 21, 2005.

Barnett GA, Sung EJ, Lin J, Hung H and Park H. 2003. An examination of Canada's inter-provincial networks. Paper presented in Sunbelt XXIII International Social Network Conference. Cancun, Mexico February, 2003.

Barnett GA. 1999. The social structure of international telecommunications. In Sawhney, H and GA Barnett. (Eds.) *Progress in communication sciences, Volume XV: Advances in telecommunications*, 151-186. Stanford, CT: Ablex.

Barnett GA & Choi Y. 1995. Physical distance and language as determinants of the international telecommunication network. *International Political Science Review, 16* (3), 249- 265.

Barnett GA & Park HW. 2005. The structure of international Internet hyperlinks and bilateral bandwidth. *Annales des Telecommunications*, 60(9-10), 1115-1132.

Brunn S & Dodge M. 2001. Mapping the "worlds" of the World-Wide Web: Restructuring global commerce through hyperlinks. *American Behavioral Scientist 44*(10), 1717-1739

Blood R. 2000. Weblogs: a history and perspective. Retrieved March 20, 2006 from: *http://www.rebeccablood.net/essays/weblog_history.html*

Cairncrass F. 1997. *The death of distance: How the communications revolution will change our lives*.Boston, MA: Harvard Business School Press

Cornfield M, Carson J, Kalis A, Simon E. (2005). Buzz, blogs, and beyond: The Internet and the national discourse in the fall of 2004. Pew Internet and American Life report. Retrieved February 26, 2006 from: *http://www.pewinternet.org/ppt/BUZZ_BLOGS__BEYOND_Final05-16-05.pdf*

DudaR, Hart P & Stork D. 2000. *Pattern Classification,* New York: John Wiley & Sons, Inc.

Fischer C. 1982. *To Dwell among Friends : Personal Networks in Town and City*. University Of Chicago Press.

Florida R. 2002. *The rise of the creative class*. NY, Basic books

Halavais A. 2000. "National borders on the world wide web. *New media & society,* 1,3, 7-28

Halavais A. 2002. Blogs and the "social weather". Conference of Internet Research 3.0. Maastricht, Netherlands.

Ihaka R and Gentleman R. 1996. R: A language for data analysis and graphics, *Journal of Computational and Graphical Statistics* **5**, 299-314.

King AD. 1991. Urbanism, Colonialism, and the World economy: Cultural and spatial foundations of the world urban system. London: Routledge.

King AD. 1990. *Global Cities: Post-imperial and the Internationalisation of London*. London: Routledge.

Kotkin J. 2001. *The new geography: how the digital revolution is reshaping the American landscape.* NY: Random House.

Kumar R, Raghavan P, Rajagopalan S, & Tomkins A. 1999. Trawling the web for emerging cyber-communities. WWW8 / *Computer Networks, 31*, 1481-1493.

Lin J & Halavais A. 2004. "Mapping the blogosphere in America. *Thirteenth International World Wide Web Conference*, New York City, May.

Lin J & Halavais A. 2005. Geographical distribution of Weblogs in the US. Paper presented to ICA general conference, *New York City, May*

Lin J & Halavais A. 2005. Comparison of political weblogs and mainstream media. Paper presented at the International Communication Association 2005 General Conference. NYC, May 2005

Logan J & Molotch H. 1987. *Urban fortunes.* Berkeley: University of California Press, 1987

Garton L, Haythornwaite C, & Wellman B. 1997. Studying online social networks. *Journal of Computer Mediated Communication: 3(1)*. Retrieved February 13, 2006, from: *http://jcmc.indiana.edu/vol3/issue1/garton.html* .

McLuhan M. 1968. *War and Peace in the Global Village*. New York: Bantam

Mitchell W. 1999. *E-topia: "Urban life, Jim-but not as we know it"*. MA: Boston, The MIT Press

Park R. 1984. The city: Suggestions for the investigation of human behavior in the urban environment. In Robert Park, Ernest Burgess (Eds.), *The city*. Chicago: The University of Chicago Press.

Park HW. 2003. Hyperlink network analysis: A new method for the study of social structure on the web. *Connections, 25*(1), 49-61.

Sassen S. 1991. *The global city*, Princeton N.J.: Princeton University Press

Wellman B. 2001. Computer networks as social networks. *Science*, 243. 2031-2034

Wellman B, Hasse A, Witte J & Hampton K. 2001. Does the Internet increase, decrease, or supplement social capital? *American Behavioral Scientist*, *45*(3), 436-455

Wellman B & Gulia M. 1999. Net surfers don't ride alone: Virtual communities. In Kollock P & Smith M. (Eds.) *Communities and cyberspace*.

Wellman B, Gulia M & Potter S. 2002. The Social Network Basis of Interpersonal Resources for Coping with Stress. In *Socioeconomic Conditions, Stress and Mental Disorders: Toward a New Synthesis of Research and Public Policy.*

Weiss M. 1989. *The clustering of America*. New York: Harper & Row Publishers.

Winer D. 2003. *What makes a weblog a weblog?* Retrieved April 20, 2004 from: *http://blogs.law.harvard.edu/whatMakesAWeblogAWeblog*

Zook MA. 2001. Being connected is a matter of geography. *netWorker, 5*(3), 13-16.

CONNECTIONS 27(2): 25-37

http://www.insna.org/Connections-Web/Volume27-2/Johnston.pdf

Modeling Indegree Centralization in NetSAS: A SAS Macro Enabling Exponential Random Graph Models

M. Francis Johnston[1]
Center for East-West Medicine; University of California Los Angeles

Xiao Chen
Academic Technology Services; University of California Los Angeles

Phillip Bonacich
Department of Sociology; University of California Los Angeles

Silvia Swigert
School of Education; University of California Los Angeles

The dual purpose of this paper is to (1) introduce SAS computer code (NetSAS) facilitating ERGM analysis of network data and (2) empirically investigate estimation and interpretation of the parameter for indegree centralization. NetSAS directly transforms square-matrix network data into rectangular-matrix dyadic data, thereby eliminating the need for computations exogenous to SAS and extensive data management. The macro is illustrated through estimation on 7 graphs of 21 nodes that vary from 0 to 100% on the conventional graph theoretic measure of indegree centralization. ERGM in a conventional statistical package may facilitate wider use of and further dialogue about the meaning, interpretation, and advancement of the ERGM framework.

INTRODUCTION

Exponential random graph modeling (ERGM, also known as p-star) is a statistical technique for modeling structural properties of networks (Snijders, Pattison, Robins, & Handcock, 2004). Wasserman and Pattison (1996) provide a rationale for modeling dyadic, triadic, subgroup, and entire network characteristics approximately via maximum pseudolikelihood (MP) methods in logistic regression (Wasserman and Pattison 1996, p. 417). Crouch and Wasserman (1998) introduce the PREPSTAR program to calculate preliminary output, along with fairly extensive code for transferring the input into and managing it in SAS. Here, we introduce NetSAS, a macro that enables statistical analysis of networks in SAS. NetSAS directly produces dyadic network data from which SAS can immediately produce basic statistics about the network and carry our ERGM.

To illustrate use of NetSAS, we engage an issue of long-standing importance in the field of network analysis -- centralization (Wasserman & Faust, 1994, pp. 175-7). For directed graphs, there are several operationalizations: indegree, outdegree, betweenness directed, closeness directed, eigenvector centrality, radiality and integration (Costenbader & Valente, 2003, p. 285). Following Crouch and Wasserman (1998), NetSAS provides the ability to model outdegree centralization and indegree centralization.

We review the graph theoretic and ERGM definitions of indegree centralization and show conceptually the issue of cross-dyadic dependency, which we illustrate with an example. We then empirically investigate the estimation and interpretation of the indegree centralization parameter on 7 graphs, each composed of 21 nodes. Empirically, our primary finding is a modest correspondence between ERGM estimation of the indegree centralization parameter and the conventional graph theoretic measure of indegree centralization. This relationship appears to be mediated somewhat by the effects of cross-dyadic dependency. By enabling analysis in a conventional statistical program, we aim to facilitate wider dialogue about the meaning, interpretation, testing, and advancement of ERGM.

---

[1] Address correspondence to: Michael Francis Johnston, Ph.D.; Assistant Researcher; UCLA Department of Medicine; Division of General Internal Medicine & Health Services Research; Center for East-West Medicine; 2428 Santa Monica Blvd., Suite 308; Santa Monica, CA 90404; Telephone: 310-453-7679; Fax: 310-315-1856. For an electronic copy of the SAS program, email Dr. Michael Johnston at Johnston@ucla.edu. We thank staff of the UCLA ATS Statistical Consulting group for providing statistical advice. We also thank Paulette Lloyd for commenting on a draft. We bear exclusive responsibility for any errors.

NetSAS

Hitherto, analysts wishing to experiment with ERGM have relied either on PREPSTAR, or highly specialized computer programs such as StOCNET and PSPAR, or even computer languages such as R. Of these ERGM-enabling options, Crouch and Wasserman (1998) created PREPSTAR to facilitate computations in SAS by using a C+ environment to calculate a range of network parameters and then providing extensive SAS code for data input, merging, management, and finally analysis. The procedure is somewhat cumbersome and the PREPSTAR algorithms are not easily interpretable to those unfamiliar with C+.

Inspired by Crouch and Wasserman, we have developed a macro we call NetSAS. NetSAS is a set of self-contained programming statements that shape conventional network data into a rectangular dyadic data matrix format that also provides a range of standard network statistics and ERGM network statistics. The data output by the macro is immediately analyzable by logistic regression in SAS. The macro is in Appendix 1 and includes some additional comments in the program itself.

NetSAS is comprised of two macro programs. The first, NetSAS Part I, produced basic network statistics. The second, NetSAS Part II creates ERGM statistics. Each macro program begins with the line "%macro" and ends with the line "%mend;". To activate the macro, simply highlight the entire macro and press run (either the SAS running person icon or the Function 3 key [F3]). To obtain results of basic network statistics, run the line "%netstat(5, d:network.txt, netstats);" where "network.txt" refers to the input data set and "netstats" refers to the output dataset. To obtain the ERGM statistics, run the line "%pstar(21, d:network.txt, tdyadic);". The macro is written with the assumption that the txt file is a square matrix located on the D drive.

NetSAS Part II outputs a SAS file titled "tdyadic", which is a rectangular-shaped dyadic data matrix composed of one row for each of the directed nodal pairs. The macro transforms the input matrix, a square gxg network matrix where g is the number of nodes, into an output matrix, a rectangular dyadic data matrix in which each dyad is one row. There are a total of (g)*(g-1) rows in the rectangular dyadic data matrix (following convention, the diagonal of the original network matrix, node-to-itself relations, is excluded). The number of dyads (rows) in a rectangular dyadic data matrix is the number of observations, for which we reserve the symbol "n". The dyadic data matrix includes column vectors for all network statistics produced in PREPSTAR: density, mutual, outstars, instars, mixed stars, transitivity, cycles, outdegree centralization (also known as degree centralization) and indegree centralization (also known as group prestige).

Once the macro has produced the dyadic data matrix, take a few moments to examine the data. One step is to examine the dyadic structure of the new dataset by printing out the nodal relations, which entails the "From" node of the directed relation, the "to" node of the directed relation, and the value of the relation (1 if there is a relation between the nodes and a 0 otherwise). SAS code to do so is provided underneath "Comment 1". A second step is to examine the network statistic values in the rows (see "Comment 2"). A third step is to examine the frequencies for variables of interest (see "Comment 3").

The next step is to fit a logistic regression model to the data (see code under "fitting the model"). When entered, the SAS code will generate output, from among which a few pieces of information are vital. Towards the top, "number of observations read" indicates the total number of directed node-to-node relationships. Further on down, under "Analysis of Maximum Likelihood Estimates," is a listing of the parameters in the model, their point estimates, standard errors, Wald Chi-Square Value, and probability of significance. Finally, there is a suite of statistical procedures for assessing model fit, which, as we describe in greater depth below, are very important in ERGM. Allison (1999) provides an excellent description of how to use SAS to carry out preliminary data characterization methods, the logistic regression procedure, and diagnose any model specification problems.

Defining Indegree Centralization: Graph Theoretic and ERGM

Indegree centralization is, roughly, a measure of the variability of actor scores on indegree centrality (Wasserman & Faust, 1994, pp. 176). When one actor's degree centrality score is high compared to the rest, the centralization score for the network as a whole will be high. Conversely, when actors have relatively equal degree centrality scores, centralization will be low. Freeman (1979) provides the conventional graph theoretic measure of indegree centralization (Formula 1). Note that indegree centralization is normalized so that scores range from 0% (a circle graph) to 100% (a star graph). In Formula 1, C_{FID} stands for a measure of centralization as defined by Freeman based upon vertex indegree, $L_{ID}(v^*)$ denotes the vertex with the largest indegree, $L_{ID}(vi)$ refers to the indegree of a vertex, and g refers to the number of vertices in the original square matrix (Wasserman & Faust, 1994, p. 180, 177).

$$C_{FID} = \left[\sum_{i=1}^{g} L_{ID}(v^*) - L_{ID}(V_i) \right] \Big/ (g-1)^2 \qquad (1)$$

In ERGM, indegree centralization and other network statistics are calculated via change score statistics. The general formula for change score statistics is Formula 2 (Anderson et al, 1999, p 48), where $z(x_{ij}^{+})$ refers to the situation in which the tie from node I to node j is forced to be present, and $z(x_{ij}^{-})$ refers to the situation in which the tie from node I to node j is forced to be absent. Formula 2 indicates that change scores are actually calculated in one of two ways: (1) Existent Relation Present – Existent Relation Hypothetically Absent, or (2) Non-Existent Relation Hypothetically Present – Non-Existent Relation Absent. Essentially, change scores measure how a particular network statistic would differ if the social network under scrutiny were to change by either the addition or subtraction of one social network tie. In the rectangular dyadic data matrix, there is one column vector for each network statistic so that the effect of adding or subtracting a tie is carried out for each dyadic relationship (that is, each row). Those readers who wish to review a detailed example of how change scores are constructed may find Crouch and Wasserman (1998) to be helpful.

$$\varpi_{ij} = \log\left\{\frac{\Pr(X_{ij}=1|X_{ij}^c)}{\Pr(X_{ij}=0|X_{ij}^c)}\right\} = \theta'[z(x_{ij}^+) - z(x_{ij}^-)] \quad (2)$$

The formula used to estimate indegree centralization is based on a measure of the number of choices received (Anderson et al., 1999, p. 57), which is a variance-based measure. In Formula 3 (Wasserman & Faust, 1994, page 180), C_{VID} is the variance-based definition of indegree centralization, $I(v_i)$ represents the indegree of the i^{th} node, $\bar{I}$ denotes the average nodal indegree.[2]

$$C_{VID} = \left[\sum_{I=1}^{g}\left(I(v_i) - \bar{I}_{ID}\right)^2\right] \Big/ (g-1) \quad (3)$$

One of the strengths of the variance-based measure of indegree centralization in comparison to the conventional graph theoretic measure of indegree centralization is that the variance-based measure allows for a larger number of change score values.[3]

Variance-Based Indegree Centralization Reveals Cross-Dyadic Dependency in ERGM

Unique to the calculation of network statistics in a change score framework is what we refer to as cross-dyadic dependency. To discuss this in depth with reference to indegree centralization, we first note that there will be, at most, "n" distinct values for the indegree centralization change scores. Consider a 10x10 square matrix will become a rectangular matrix consisting of 90 rows. For such a matrix, there are [g*(g-1) = 10*9 =] 90 dyadic relations. If the dyads were completely independent of each other, there would potentially be 90 distinct values for the indegree centralization change scores.

Even with independence, there might be less than 90 distinct values for the indegree centralization change scores. One reason is very common, namely that in any dataset some values might occur more than once. Imagine that final grades for a class of 90 undergraduate students could potentially range from 0 to 100 total possible points. In this individualistic example, undergraduates would be considered as independent of each other but it is likely that a few might have the same number of total points. Despite independence among observations in this example, there would be less than 90 distinct values for the final numeric grade. The network equivalent of this individualistic example is to note that some vertices might have the same indegrees, which would result in a fewer number of indegree centralization change scores than the possible maximum. This is not what we mean by cross-dyadic dependency.

By cross-dyadic dependency, we are referring to the realization of a much fewer number of values for indegree centralization change scores (and other network statistics) than the maximum possible because of dependencies among the dyads which arise because individual vertices are involved in more than dyad. This becomes obvious when the rectangular matrix of dyads is arranged by the "to" vertices. For example, consider output from an analysis of a size 10 network from the Knoke bureaucracies in UCINET, the matrix titled Money. Table 1 shows all of the node-to- node relations that involve Node 5 (indegree=1) and Node 8 (indegree=6). Node 5 only receives money from one organization, Node 1, which is reflected in the column labeled Y. There is only a single 1 which is located in the first row-the row that corresponds to the directed relationship FROM node 1 TO Node 5. Since Node 5 does not receive money from any of the other organizations, all of the other rows have a 0 in the column labeled Y. In contrast, Node 8 receives money from six other organizations.

Table 1: Indegree Centralization Scores (Variance)

FROM Node	TO Node	Y	Change Score CID
1	5	1	-0.36667
2	5	0	-0.16667
3	5	0	-0.16667
4	5	0	-0.16667
6	5	0	-0.16667
7	5	0	-0.16667
8	5	0	-0.16667
9	5	0	-0.16667
10	5	0	-0.16667
1	8	1	0.74444
3	8	1	0.74444
4	8	1	0.74444
5	8	1	0.74444
7	8	1	0.74444
9	8	1	0.74444
2	8	0	0.94444
6	8	0	0.94444
10	8	0	0.94444

Cross-dyadic dependency arises from calculating indegree centralization by applying a variance-based operationalization

[2] Although Wasserman and Faust use "g" as the denominator, we use (g-1) for the sake of consistency with the conventional way of computing variance.

[3] The conventional formula does not distinguish differences between the nodes in terms of indegree centralization, and, as a result, a large number of node-node relations will cluster into an insufficient number of categories to employ the resulting vector as a variable in a logistic regression analysis.

within a change score procedure. Note that all dyads which involve Node 5 as the "to" node has either one of two values for the indegree centralization change score, either -0.36667 or -0.16667. Note furthermore the pattern organizing these realizations. All dyadic relations involving Node 5 as the "to" when the tie is actually existent in the data (Y=1) have an indegree centralization change score of -0.36667. When the tie is actually non-existent in the network (Y=0), the indegree centralization change score is -0.16667. This pattern also holds for all couples involving Node 8 as the "TO" node (0.7444 when Y=0 or 0.94444 when Y=1) and each of the other nodes.

If the dyads were independent of each other, there could be as many as 90 distinct indegree centralization change score values. Because of cross-dyadic dependency, however, these 90 dyadic relations would fall into at most 2g = 2*10 = 20 indegree centralization scores. As discussed above, calculating indegree centralization by applying a variance-based operationalization within a change score procedure for matrices of size 10 will oftentimes result in less than 20 indegree centralization scores, because nodes with the same indegree will have the same value for their indegree centralization change score.

Table 2. 13 Categories of Indegree Centralization Scores

FROM Node	TO Node	Y	Change Score CID	# node-node relations
10	6	0	-0.38889	30
4	7	1	-0.36667	2
10	7	0	-0.16667	16
8	10	1	-0.14444	2
2	10	0	0.05556	7
7	2	1	0.07778	3
10	2	0	0.27778	6
8	9	1	0.30000	4
10	9	0	0.50000	5
9	3	1	0.52222	5
10	3	0	0.72222	4
9	8	1	0.74444	6
10	8	0	0.94444	3

Variance-Based Indegree Centralization in ERGM

Of the possible 20 indegree centralization change score values, there are only 13 in Money. Each realization corresponds to a particular kind of node-node relation that is based upon the "to" node and the value of "Y" (see Table 2). Notice that the most negative indegree centralization change score category is -0.38889, which involve nodes with a zero indegree as the "to" node. This signifies that if a tie were to be added to a node with zero indegree, there would be a decrease in the amount of indegree centralization in Money. The next smallest change score category, -0.36667, involves a node that has an indegree of 1. This signifies that if a tie were to be eliminated to a node with indegree of one, there would be a decrease in the amount of indegree centralization in Money. The third smallest change score category, -0. 16667, shows that if a tie were to be added to a node with indegree one, there would be a decrease in the amount of indegree centralization in Money.[4] Informing the calculation of change scores for indegree centralization is the general idea that if all the nodes had exactly the same indegree, the graph would be entirely non-centralized.

The two largest change score values are associated with the node with the largest indegree, Node 8: 0.94444 and 0.74444. The largest occurs when Node 8 is changed from a node with indegree of six to a node with indegree of seven, thereby increasing the amount of indegree centralization in the graph, even compared to the second largest which occurs when Node 8 is changed from a node with indegree of six to a node with indegree of five. Table 2 shows one of the desirable properties of using the variance-based measure of indegree centralization to calculate change scores, namely that when the node-node relations are ordered by magnitude of the change score values, the dyadic relations with the largest indegrees score the highest.

Estimating Indegree Centralization in ERGM

In the previous section, we suggested that the method of calculating change scores, though it may account for the non-independence among dyads, also brings about cross-dyadic dependency. Specifically, we showed that those node-node observations with the same "to" node will have either one or two values for the change score of indegree centralization. Recall that a primary assumption of generalized linear models, of which logistic model is a specific example, is that observations are independent of each other (Agresti, 2002, p. 116, 455).[5] What is the impact of violating this assumption of statistical independence?

A first order of concern prompts the question: Does cross-dyadic dependency bias the coefficient estimate for indegree centralization? One way to approach this question is to conceptualize cross-dyadic dependency as a type of clustering similar to students nested in a classroom — dyadic relations with the same "to" node can be grouped together as being part of the same setting. In this way, those who take a standard approach to statistical modeling would seem to argue no, the coefficient

[4] We find this negative value to be mildly counter-intuitive. We had expected that taking away a tie to a node with one indegree would increase the amount of indegree centralization. However, we do not consider this to be strongly counter-intuitive because the decrease in indegree centralization is much greater when a tie is taken away from a one-degree node than when a tie is added.

[5] See also Hardin & Hilbert, 2003, p. vii : "…[B]eing likelihood based, [Generalized Linear Models] assume that individual rows in the data are independent from one another. However, in the case of longitudinal and clustered data, this assumption may fail. The data are correlated."

estimate is not biased.[6] We hasten to add, however, that this issue is now being debated in a large and rapidly growing area of statistical literature addressing what is variously labeled as cluster-level covariates, correlated binary data, or random effects modeling. In this area, some statisticians advocate for a more complicated model that includes a cluster-specific random effect term within the logit model (for a discussion, see Hosmer and Lemeshow 2000, pp. 308-330). Beyond the scope of our paper is another special branch of statistical modeling known as Generalized Estimating Equation (GEE), which adjusts both parameter estimates and standard errors for clustering by using a population average model (Hardin & Hilbert, 2003). Both random effects and GEE may provide much traction for modeling correlated binary data. But they are still relatively new areas of research, and many modeling details are in the process of being worked out. After reviewing much of this research, Hosmer and Lemeshow (2000, p. 327) write: "we think it best to proceed cautiously when fitting cluster-specific models."

A second order of concern prompts the question: Does clustering affect the standard error estimate for indegree centralization? The answer appears to be yes. From the perspective of those utilizing a conventional logistic regression modeling framework, when clustering impacts variance, it will almost always inflate the variance of the binomial response variable and only rarely in practice deflates the variance (Collett, 2003, p. 195). Various models have been proposed to weigh the data to compensate for inflated variance (Collett, 2003, pp. 202-213). Since variance is an important component in the calculation of standard errors in logistic regression (see Collett 2003, Chapter 3 for details), it is likely that problems with the variance would lead to bias in the standard errors for indegree centralization. This might be the factor that motivated Wasserman and Pattison (1996, p. 415, 424) to advocate for testing overall model fit (by comparing model fit with and without the parameter) instead of examining inferential tests for particular parameters in their original p-star paper.

More recently, Snijders and colleagues (2004, p. 7) have claimed that the chi-squared likelihood ratio tests, which logistic regression packages automatically compute to evaluate the statistical significance of particular coefficient parameters, are problematic.[7]

We summarize our understanding of parameter estimation for indegree centralization with the following five points.[7]

1. Conceptually, the parameter estimates the extent to which indegree centralization contributes to a graph's overall structure by computing the extent to which "the actual network" differs from "the set of all hypothetical networks distinguished by just a one tie."
2. Computationally, the indegree centralization parameter is estimated in a change score format with a variance-based operationalization.
3. Because of cross-dyadic dependency in the data, observations with the same "to" node will have at most two distinct values for the indegree centralization change score.
4. Cross-dyadic dependency may bias the coefficient estimate for indegree centralization, but this point is debated.
5. Cross-dyadic dependency likely biases estimation of standard error.

We now turn to empirically examine the estimation of the indegree centralization parameter. To maximize insight into the basic workings of inferential statistics in ERGM, and avoid the issue of biased standard errors, we carry out this work out in a bivariate framework, where testing a coefficient parameter is equivalent to testing overall model fit (Hays, 1963, pp. 354, 375, 465).

Data and Analysis

In this section, we begin the process of testing parameter estimation of indegree centralization in the ERGM framework with selected graphs that have twenty-one nodes. We choose to start with networks of size 21 for two primary reasons. First, this is a network size of interest to those who carry out research in education in that many classrooms have approximately 20 students, as is the case for data analyzed in Anderson et al. (1999, pp. 42-44). Second, there is well-known data available with 21 nodes (Krackhardt, 1987). The first graph we choose to examine is Circle, in which each node chooses two others. Circle is considered the most non-centralized, or most egalitarian, of graph structures. On the other side of the spectrum, we have chosen Hierarchy, a graph in which one node receives ties from each of the other 20 nodes but this node does not choose the other nodes and the other nodes do not select each other (in other words, this is a directed star graph). Additionally, we analyze three well-known graphs collected by Krackhardt (1987) concerning relations between 21 managers in a company, manufacturing high-tech equipment on the west coast of the United States. Each manager was asked two questions. Answers to the first question ("To whom do you go to for advice?") are recorded in a graph we label as "Advice."

[6] For example, Long (1997, p. 50), after a mathematical proof specifically on the impact of clustering on coefficient estimation writes: "Consequently, the probability of an event is unaffected by the identifying assumption regarding Var ($\mathcal{E}$ | x) . While the specific value assumed for Var ($\mathcal{E}$ | x) is arbitrary and affects the β's, it does not affect the quantity that is of fundamental interest, namely, the probability that an event occurred…The critical point is that while the β's are not affected by the arbitrary scale assumed for $\mathcal{E}$, the probabilities are not affected. Consequently, these probabilities can be interpreted without concern about the arbitrary assumption that is made to identify the model. That is to say, the probabilities are estimable functions. Further, any function of the probabilities is also estimable. Importantly, we can interpret changes in probabilities and odds, which are ratios of probabilities."

[7] "To estimate the parameters, the pseudo-likelihood method continued to be used, although it was acknowledged that the usual chi-squared likelihood ratio tests were not warranted here…" (Snijders et al., 2004, p. 7).

Information from the second question ("Who is your friend?") is in the overall graph, "Friendship." Also, collected from company documents was information about a third type of tie: "To whom do you report?" We label this overall graph as "Reports." We do not provide graphics for Circle, Hierarchy, Advice, and Reports because these networks are very straightforward.

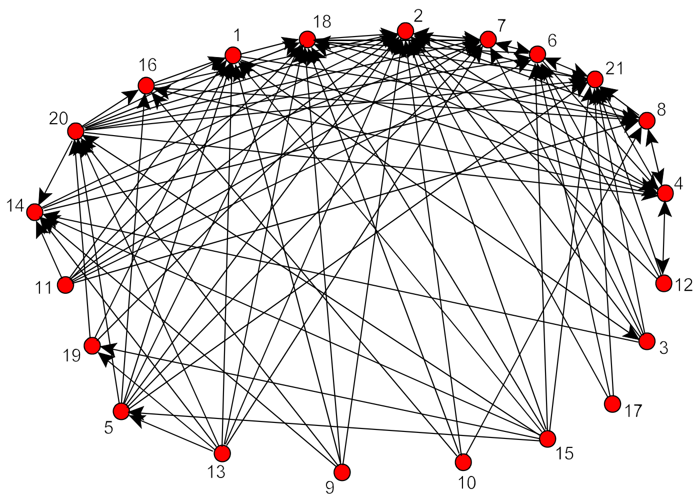


Figure 1. Graphic for Actor 2 (Advice)

Krackhardt also asked each of the managers to indicate what he or she perceived to be the relations among all other managers. So, for each actor, there is a graph for advice relations among the 21 actors and a graph for the friendship relations among each of the 21 actors. From these 42 matrices, we selected the perception of the second actor of the advice relation among the 21 managers, because it has a relatively high amount of indegree centralization (see Figure 1). We also selected the perception of Actor4 of the friendship relations among the 21 managers, because it has a low amount of indegree centralization (see Figure 2).

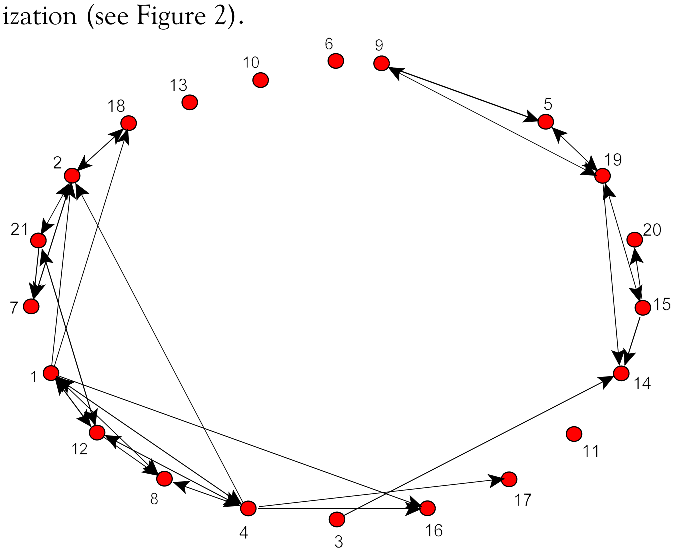


Figure 2. Graphic for Actor 4 (Friendship)

We begin by identifying the amount of indegree centralization in each of the networks. To provide a sense for the level of centralization in these networks, we first compute indegree centralization scores in UCINET (Borgatti, Everett, & Freeman, 1999). Other measures of centralization could have been used, for example a variance-based measure of centralization. However, we chose the standard calculation (graph indegree centralization) because it is widely used and recognized.[8] Scores are shown in Table 3. Ordered from most to least centralized, the graphs are: Hierarchy, Actor2, Advice, Reports, Friendship, Actor4, and then Circle.

Table 3 also contains coefficient and standard error estimates for the indegree centralization parameter as computed in SAS. First, note that the parameter estimates for Hierarchy and Circle are both very high and in the expected direction: Hierarchy is highly positive and Circle is highly negative. But, corresponding standard errors are also extremely high, and therefore the p-values show the structure to be insignificant, though in truth the structure is very significant. The standard errors are inflated because the logistic regression model is being fit to data with a very small number of change score values.

Next, we turn to examine Actor2, which is a relatively centralized graph. As expected, there is a relatively strong important coefficient value and a low standard error. Moreover, the chi-square statistic identifies the amount of centralization in the graph as statistically significant.

Consider now the graph for Advice, which is a less centralized graph. The estimated coefficient is smaller (1.663, compared to 2.820 for Actor2). Also, the estimated standard error is small, so that the p-value is statistically significant ($p<.0001$). Surprisingly, Reports has a much higher coefficient value than is the case for Actor2 and Advice. There is also a much higher standard error. In part, we attribute this higher standard error to sparsity of ties in the matrix: there are only 20 ties present from a total possible of 420 and only 9 different indegree centralization change score values, which are skewed right.

The next most centralized graph is Friendship. Though similar in centralization score to Reports, Friendship has many more ties (102 vs. 20) and more change score values (16 vs. 9). Moreover, the structure of relations is less skewed: two nodes each have an indegree of 1, 2, and 3. Three nodes have an indegree of 4, five nodes an indegree of 5, and four an indegree of 6. At the other tail, two nodes have an indegree of 8 and one an indegree of 10. The parameter estimate is smaller than the estimates for Actor2, Advice, and Reports (0.730 vs. 2.82, 1.663, 5.335). Note, however, in contrast to the previous matrices, that the p-value is non-significant.

Finally, we turn to the graph for Actor4, which has the second lowest indegree centralization score. There are a slightly higher number of ties compared to Reports (36 vs. 20). Yet, more ties

[8] Additionally, the variance measure of centralization is not very standardized. Suppose, for example, that we examine the probability (p) that any pair of individuals is connected in a random graph. The variance indegree will be Np(1-p), the variance of the binomial distribution where N is the number of alternative partners for each person (one less than the group size).

does not translate into more spread—the number of nodes with distinct indegree scores is the same (5), as is the number of distinct change score values (9). Indeed, the nodal indegree ranges from 0-4 for Actor4, as compared to 0-7 for Reports, which suggests that the change score values will also have a much more restricted range. One quality of the graph for Actor4, when compared to Reports, is that it is less skewed: five nodes have an indegree of 0, four with an indegree of 1, five with an indegree of 2, six with an indegree of 3, and one with an indegree of 4. All the nodes fall close around the average indegree of 1.7. The parameter coefficient is very low (0.195), indicating a very small slope. Also, the standard error is a bit high (1.441), perhaps because of the lack of spread in nodal indegrees. Similar to Actor4, the p-value is non-significant (0.8924).

Table 3. Indegree Centralization Network Statistics, Graph and ERGM Parameterization

Graph	Total Number of Ties	Number of In-degree Values	Number of Change Score Values	Graph Indegree Central-ization	P-star coef.	Standard Error	P-value
Hierarchy	20	2	2	100.00%	18.941	31568.400	<0.9995
Actor2	110	14	24	77.50%	2.820	0.310	<0.0001
Advice	190	10	20	47.00%	1.663	0.284	<0.0001
Reports	20	5	9	31.75%	5.335	0.929	<0.0001
Friendship	102	8	16	27.00%	0.730	0.540	<0.1764
Actor4	36	5	9	12.00%	0.195	1.441	<0.8924
Circle	42	1	2	0.00%	-264.30	2656.900	<0.9208

DISCUSSION

This paper introduces NetSAS, a SAS macro that transforms conventional network data into dyadic data and provides a set of conventional network statistics, both of which facilitate ERGM in SAS. We use the program as a launching point to examine estimation of the indegree centralization parameter in ERGM. We identify and discuss the origin and some consequences of cross-dyadic dependency. We further suggest that the idea of clustering may be one fruitful way to conceptualize cross-dyadic dependency.

We agree with Seary and Richards (2000, p. 87) that it is desirable to proceed with caution when estimating indegree centralization in an ERGM format, though we are by no means calling for researchers to abandon a variance-based operationalization. When we empirically tested parameterization by analyzing seven matrices that span the range of indegree centralization according to the conventional graph theoretic measure, we find that the ERGM framework is generally able to identify those networks with more than a modest amount of indegree centralization. We recognize that our research design is insufficiently rigorous to claim that a variance-based measure of indegree centralization is justifiable in the ERGM framework. However, we claim that these results provide a warrant for further research on this topic, especially that which examines the usefulness of conventional statistical procedures in correcting for clustering. One traditional approach would be to apply a post-hoc overall adjustment to standard errors through a technique often referred to as the sandwich variance estimator (Hardin and Hilbe 2003, p. 5). Another traditional technique, known as fixed effects modeling, would be to include an indicator variable for each node, omitting one from the model. Multilevel modeling would enable adjustments to both the parameters and standard errors in light of clustering. Yet another technique is the method of population average known as GEE (Hardin and Hilberg 2003).

Our paper raises another issue worthy of further attention, namely the relation between the ERGM parameter estimate of indegree centralization and the graph theoretic measure. The analyses presented here suggest correspondence, but many details remain to be worked out. For example, what does it mean to hypothesize that the null value of the coefficient is zero? Is this equivalent to hypothesizing that the variance-based graph measure of indegree centralization is 50%? Will the expected value of the distribution vary substantially by graph size? Recent research by Tallberg (2004) suggest possible ways to address these and related questions about model testing using simulation methods.

Efforts to test and compare measures within and between data-sets (e.g. Costenbader and Valente 2003) provide a scientific foundation fostering the diffusion of this statistical network methodology. Programs permitting ERGM in conventional statistical packages are critical for enabling a larger number of people to participate in building a more practical foundation with well-understood strengths and limitations. In this paper, we have studied indegree centralization, showing how further scrutiny may uncover issues worthy of further attention. More widespread participation in dialogue about the meaning, interpretation, and testing of ERGM is critically important for further advancing and diffusing this network science innovation.

References

Agresti A. 2002. *Categorical data analysis.* Hoboken, NJ: Wiley-Interscience.

Aldrich JH & Nelson FD. 1984. *Linear probability, logit, and probit models.* Thousand Oaks, CA: Sage Publications.

Allison PD. 1999. *Logistic regression using the SAS system: Theory and application*. SAS Press and John Wiley Sons.

Anderson C, Wasserman S, & Crouch B. 1999. A p* primer: Logit models for social networks. *Social Networks,* 21: 37-66.

Borgatti S, Everett M, & Freeman L. 1999. *UCINET 5 for Windows: Software for social network analysis.* Natick: Analytic Technologies. http://www.analytictech.com/ucinet_5_description. htm

Collett D. 2003. *Modeling binary data, second edition.* New York: Chapman and Hall/CRC.

Cook RD & Weisberg S. 1999. *Applied regression including computing and graphics.* New York: Wiley-Interscience.

Costenbader E & Valente T. 2003. The stability of centrality measures when networks are sampled. *Social Networks,* 25: 283-307.

Crouch B & Wasserman S. 1998. A practical guide to fitting p* social network models. *Connections,* 31: 87-101.

Frank K, & Yasumoto J. 1998. Linking action to social structure within a system: social capital within and between subgroups. *American Journal of Sociology,* 104: 642-86.

Freeman L. 1979. Centrality in social networks: conceptual clarification. *Social Networks,* 1: 215-239.

Hardin JW & Hilbert JM. 2003. *Generalized estimating equations.* New York: Chapman & Hall/CRC.

Hays WH. 1963. *Statistics for psychologists.* New York: Holt, Reinhart, and Winston.

Holland PW & Leinhardt S. 1975. The statistical analysis of local structure in social networks. *Sociological Methodology, 197*: 1-45.

Hosmer D & Lemeshow S. 1989. *Applied logistic regression.* New York: Wiley-Interscience.

Hosmer D & Lemeshow S. 2000. *Applied logistic regression* (2nd ed.). New York: Wiley-Interscience.

Koehly L & Wasserman S. 1996. Classification of actors in a social network based on stochastic centrality and prestige. *Journal of Quantitative Anthropology,* 6: 75-99.

Knoke D & Wood J. 1981. *Organized for action: Commitment in voluntary associations.* New Brunswick, NJ: Rutgers University Press.

Knoke D & Kuklinski J. 1982. *Network analysis*. Thousand Oaks, CA: Sage Publications.

Krackhardt D. 1987. Cognitive social structures. *Social Networks,* 9: 104-134.

Lazega E & Pattison PE. 1999. Multiplexity, generalized exchange and cooperation in organizations: A case study. *Social Networks,* 21: 67-90.

Lazega E & van Duijn M. 1997. Position in formal structure, personal characteristics and choices of advisors in a law firm: A logistic regression model for dyadic network data. *Social Networks,* 19: 375-397.

Long J. 1997. *Regression models for categorical and limited dependent variables.* Thousand Oaks, CA: Sage Publications.

Lubbers M. 2003. Group composition and network structure in school classes: A multilevel application of the p* model. *Social Networks,* 25: 309-332.

Seary AJ & Richards WD. 2000. Fitting to p* models in Multinet. *Connections,* 23(1): 84-101.

Skvoretz J & Faust F. 1999. Logit models for affiliation networks. In Sobel ME & Becker MP, (Eds.), *Sociological Methodology* (253-280). Cambridge, MA: Basil Blackwell.

Snijders TAB, Pattison PE, Robins GL, & Handcock MS. 2004. New specifications for exponential random graph models. Manuscript.

Tallberg C. 2004. Testing centralization in random graphs. *Social Networks,* 26(3): 189-288.

Wasserman S & Faust K. 1994. *Social network analysis: Methods and applications.* Cambridge: Cambridge University Press.

Wasserman S & Pattison PE. 1996. Logit models and logistic regressions for social networks: I. An introduction to markov graphs and p*. *Psychometrica,* 61: 401-425.

Appendix 1: NetSAS

```
/*****************************************************
 NetSAS, Part I - produces basic network statistics
 Version 1.0 Modified August 26, 2005
*****************************************************/

/*****************************************************
 netstat program takes three parameters:
 nnodes:  number of nodes
 infile:  the input file for a network data
 outdata: name for the output sas data file
*****************************************************/

%macro netstat(nnodes, infile, outtable);
 data   indata  ;
      infile "&infile";
      input a1 - a&nnodes;
 run;

proc iml;
 use   indata  ;
 read all into x;          /*read in the sociomatrix*/
 G = nrow(x);              /*number of nodes, which is also the number of rows*/
 g2 = g*g;                 /*g2 is used for computation purposes*/
 N = g2 - g;               /*number of observations, dyadic pairs, is (g^2 - g)*/

 /*number of edges is the sum of all edges in the matrix*/
  L = sum(x);

 /*density is number of edges divided by number of dyadic pairs*/
  D = L/N;

 /*mean indegree is number of edges divided by number of nodes*/
  Mean_Indeg = L/G;

 /*mean outdegree is also number of edges divided by number of nodes*/
  Mean_Outdeg = L/G;

 /*An outstar is the number of nodes that connect outwards to exactly two nodes*/
  Stars_out = (sum(t(x)*x) -trace(t(x)*x))/2;

 /*An instar is a nodes that receives connections from exactly two nodes*/
  Stars_in = (sum(x*t(x)) - trace(x*t(x)))/2;

 /*number of nodes that have an outward connection to exactly one node
  and an inward connections from exactly one node*/
  Stars_mixed = sum(x*x) - trace(x*x);

 /*number of triads out of all possible triads with a transativity*/
      Trans_triads = trace(x*x*t(x));

 /*number of triads out of all possible triads with a cycle*/
  Cyclicity = trace(x*x*x)/3;

 /*number of dyads out of all possible dyads that have a reciprocated relation*/
  Mutual_dyad =sum(x#x)/2;

  t1 = j(1, g, 0);
      t2 = j(1, g, 0);
         do k = 1 to g;
```

```
           t1[1, k] = (sum(x[1:g, k]) - mean  indeg)**2;
        t2[1, k] = (sum(x[k, 1:g]) - mean_outdeg)**2;
     end;
   cen = sum(t1)/(g-1);
   pre = sum(t2)/(g-1);

  c = j(14, 1, 0);

  c[1, 1] = G;
  c[2, 1] = N;
  c[3, 1] = L;
  c[4, 1] = D;
  c[5, 1] = Mean  Indeg;
  c[6, 1] = Mean  Outdeg;
  c[7, 1] = Stars  out;
  c[8, 1] = stars  in;
  c[9, 1] = stars  mixed;
  c[10, 1] = trans  triads;
  c[11, 1] = cyclicity;
  c[12, 1] = mutual_dyad;
  c[13, 1] = cen;
  c[14, 1] = pre;
  names={G N L D Mean_indeg Mean_Outdeg Stars_out Stars_in Stars_mixed Transitivity Cyclicity Mutual ind_cen grp_pres};
  heading = {N STAT};
  print c [rowname=names colname=heading];
quit;
ods output c = &outtable;
%mend;
%netstat(5, d:\network.txt, netstats);

/****************************************************
  NetSAS, Part II - produces ERGM statistics
  Version 1.0 Modified September 16, 2005

  Formulae are almost entirely derived from Table 4 (page 46) from
  "A p* primer: logit models for social networks",
  Social Networks, 21(1999) 37-66.

  Expressed in matrices:

  dyadic:
  mutual:        sum(A#A')/2

  triadic
  2-out-stars:   (sum(A'*A) -trace(A'*A))/2
  2-in-stars:    (sum(A*A') -trace(A*A'))/2
  2-mixed-stars: sum(A*A) - trace(A*A)
  transitivity:  trace(A*A*A')
  cyclicity:     trace(A*A*A)/3

  average indegree per node: L= sum(A)/dim(A);
  degree centralization:              (sum(in  i - L) ^ 2) /(g-1)
                             with in  i = sum(x[1:g, i])
  group prestige:                     (sum(out  i - L) ^ 2)/(g-1)
                             with out_i = sum(x[i, 1:g])

  Formulae for change scores are derived from them.

****************************************************/
```

```
/*************************************************
  pstar program takes three parameters:
  nnodes:  number of nodes
  infile:  the input file for a network data
  outdata: name for the output sas data file

  Output data file contains change statistics on each dyad. Below is the
  description of each variable in it. Change score by definition is the
  difference between the statistic when the tie is present and the
` statistic when the tie is missing.

  Each row (dyad) is indexed by variable From and to, indicating a dyad
  (i, j).

  Variables created in outdata set are:

  Dyad-level variables:

  var1:  From           -- from the ith subject
  var2:  to             -- to the jth subject
  var3:  y              -- x[i,j], the link indicator between ith subject and jth subject
  var4:  density        -- currently defined as 1
  var5:  mutual         -- x[j,i] (rho)
                        when x[j,i] = 0, it will not be a mutual dyad,
                        so the change score = (0-0) = 0 = x[j,i]
                        when x[j,i] = 1, it wil be a mutual dyad when the tie (x[i,j]) is present
                        and not a mutual when the tie is missing so the difference is 1 = x[j,i].

Triad-level variables:

var6:  outs             -- 2-out-stars (sigma  o)
                        number of 2-out-stars when the tie is present
                        minus number of 2-out-stars when the tie is missing.

var7:  ins               -- 2-in-stars (sigma_i)
                        number of 2-in-stars when the tie is present
                        minus number of 2-in-stars when the tie is missing.

var8:  mixs             -- 2-mixed-stars (sigma _m)
                        number of 2-mixed-stars when the tie is present
                        minus number of 2-mixed stars when the tie is missing.


var9:  trans            -- transitivity (tau_t)
                        transitivity when the tie is present
                        minus transitivity when the tie is missing.

var10: cyclic           -- cyclicity (tau_c)
                        cyclicity when the tie is present
                        minus cyclicity when the tie is missing

var11: degree_cen       -- degree centralization
                        indegree centralization when the tie is present
                        minus indegree centralization when the tie is missing

var12: group  prestige  -- group presitge
                        group prestige when the tie is present
                        minus group prestige when the tie is missing

*************************************************/
```

```
%macro pstar(nnodes, infile, outdata);
 data   indata  ;
       infile "&infile";
       input a1 - a&nnodes;
 run;

proc iml;
 use   indata  ;
 read all into x;           /* read in the sociomatrix */
 g = nrow(x);               /* number of nodes */
 g2 = g*g;                  /* number of pairs = g2 - g*/

 t1 = j(1,g, 0);
 t2 = j(1,g, 0);

 /**************************************************
   calcuating change statistics, using matrix calculation
 **************************************************/

   c = j(g2, 12, 0);        /*creating a matrix for all pairs*/

       do i = 1 to g by 1;
       do j = 1 to g by 1;
            tmp = x;
            tmp[i, j] = (x[i,j]=0);
         if i ^=j then do;

         out   = (sum(t(x)*x)-trace(t(x)*x))/2 - (sum(t(tmp)*tmp)-trace(t(tmp)*tmp))/2;
         in    = (sum(x*t(x))-trace(x*t(x)))/2 - (sum(tmp*t(tmp))-trace(tmp*t(tmp)))/2;
         mixed = sum(x*x) - trace(x*x) - sum(tmp*tmp) + trace(tmp*tmp);
         trans = trace(x*x*t(x)) - trace(tmp*tmp*t(tmp));
         cyc   = trace(x*x*x)/3 - trace(tmp*tmp*tmp)/3;

         lx    = sum(x)/g;
         ltmp  = sum(tmp)/g;

              do k = 1 to g;
              t1[1, k] = (sum(x[1:g, k]) - lx)**2;
          t2[1, k] = (sum(tmp[1:g, k]) - ltmp)**2;
         end;
         grpp = sum(t1)/(g-1);
         grpm = sum(t2)/(g-1);
              grp = (-1)**(1-x[i,j])*grpp +(-1)**(x[i,j])*grpm;

              do k = 1 to g;
              t1[1, k] = (sum(x[k, 1:g]) - lx)**2;
          t2[1, k] = (sum(tmp[k, 1:g]) - ltmp)**2;
         end;
         indp = sum(t1)/(g-1);
         indm = sum(t2)/(g-1);
              ind = (-1)**(1-x[i,j])*indp +(-1)**(x[i,j])*indm;

    c[j + g*(i-1),  1] = i;                  /*from   */
        c[j + g*(i-1),  2] = j;              /*to     */
        c[j + g*(i-1),  3] = x[i,j];
        c[j + g*(i-1),  4] = 1;              /*density*/
        c[j + g*(i-1),  5] = x[j,i];         /*mutual*/
    c[j + g*(i-1),  6] = abs(out);
        c[j + g*(i-1),  7] = abs(in);
    c[j + g*(i-1),  8] = abs(mixed);
```

```
          c[j + g*(i-1),  9] = abs(trans);
          c[j + g*(i-1), 10] = abs(cyc);
          c[j + g*(i-1), 11] = ind;
          c[j + g*(i-1), 12] = grp;
        end;
  end;
end;
 create &outdata  var {From to y density mutual outs ins mixs trans cyclic
                     degree_cen group_prestige};
         append from c;
quit;
data &outdata;
  set &outdata;
  if from ~= to;
run;
%mend;
%pstar(5, d:\network.txt, tdyadic);
options nocenter nodate;


********************************;
*Becoming familiar with the data;
********************************;

*Coment 1. Examine basic data - two nodes and the relation between them;
proc print data = tdyadic;
var from to y;
run;

* Comment 2: Examine network statistics;
* reminder - degree centrality is also known as outdegree centralization;
* reminder - group prestige is also know as indegree centralization;

proc print data = tdyadic;
var density mutual outs ins mixs trans cyclic degree_cen group_prestige;
run;

* Comment 3: We examine indegree centralization more carefully;
proc freq data = tdyadic;
  tables group_prestige;
run;

********************************;
* Fitting the Model;
********************************;

* We use a single parameter, indegree centralization;

proc logistic data=tdyadic descending;
  model y = group_prestige / lackfit rsq ctable;
  output out=opred1 prob=phat;
run;
```

CONNECTIONS 27(2): 39-52

http://www.insna.org/Connections-Web/Volume27-2/5.Petróczi.pdf

Measuring tie-strength in virtual social networks

Andrea Petróczi
Kingston University, Kingston-upon-Thames, United Kingdom

Tamás Nepusz and Fülöp Bazsó
Hungarian Academy of Sciences, Budapest, Hungary

Tie-strength has been in the focus of social science research for decades, yet the use of measurement tools or scales has been relatively scarce. The aim of this study was to fill the gap and provide a tool that is able to provide a quantitative and continuous measure of tie strength in social networks. The focus was on virtual communities because the fast expansion of Internet use and the constant growth of on-line communities provide today's researchers with an excellent opportunity for effective and speedy data collection regarding tie-strength measures in these virtual social groups. The Virtual Tie-Strength Scale (VTS-Scale) consist of 11 questions and it was developed on a sample of 56 people (3080 asymmetric ties) and tested for reliability of smaller sample of 16 (204 asymmetric ties) independent sample participation regularly on a Hungarian discussion board like forum. Reliability coefficients were reassuringly high for both samples, Cronbach alphas of 0.92 and 0.86, respectively. Data triangulation offered evidence for scale validity. In summary, the VTS-Scale and its scoring method seem to provide a valid and reliable measure of tie strength in virtual communities. Although the aim of the research was to develop a tool that measures tie-strength in virtual communities, the tool can be easily modified for off-line social groups. The VTS-Scale is also capable of distinguishing between two components of tie-strength: acquaintances and friendship. However, the content of each component needs further investigation.

INTRODUCTION

Social networks - real or virtual - are collections of human communities. There are several studies (e.g., Burt, 1995; Granovetter, 1973, 1982) that examined real world/off-line social groups and have influenced our thinking about social constructions. However, these empirical studies were based on limited samples insufficient for rigorous and decisive mathematical and statistical analysis. In this article, first we give an overview of computer mediated communication and virtual communities. As the notion of tie strength in social networks is in the centre of our paper, we discuss it in details. Then methodology is described, a scale measuring acquaintance and friendship in virtual communities is introduced and statistical analysis is performed. Our discussion reflects on the validity and reliability of the VTS-Scale and points to future research.

Computer-mediated social networks

Observing and analysing on-line social networks has undeniable advantages over the face-to-face methods. Computer mediated communication (CMC) allows the researcher to overcome difficulties presented from time and distance barriers in face-to-face research (Mann and Stewart, 2000). Jones (1999), however, issues a warning not to be misled by the seductively easy access to large and textually rich data. Yet, in this research, we will use the assumptions made by Haythornthwaite (2002, p.388): "[C]haracteristics of ties hold in the mediated environments as they do in off-line environments", and (Ibid, p. 388) "on-line exchanges are as real in terms of their impact on the tie as are off-line exchanges". Wellman's earlier work (2001) has also come to the conclusion that computerised networks are, indeed, social networks.

A sense of community exists in the mind of the participants. Virtual groups of people are invisible, nevertheless off-line communities, which are given meaning by their participants. In other words, a community is as its members define it for themselves. Both offline and on-line social networks can be described by 1) their participants, 2) the content, direction, strength of their relations and ties, 3) their composition, derived from the social attributes of the participants, and 4) their complexity, which indicates the number of relations in a tie (Garton, Haythornthwaite, Wellman, 1997). Traditionally, a community has been defined by shared space and common value system (see Jones, 1997). Although it is yet to be decisively proven, it is believed that virtual communities mirror those in the 'real world' in many ways: cyber communities also share values, beliefs, norms and expectations regarding the appropriate behaviour and have a sense of identity, commitments and association (Preece, 2000).

Computer Mediated Communication (CMC) has been the focus of much research in the past decade. CMC services on the Internet range from the World Wide Web (WWW), electronic mails, mailing lists, usenet newsgroups, focus groups facilities, chats, multi-user text-based role-playing environments (MUDs), multimedia environments and conferenc ing to message boards and Internet forums (Hine, 2000; Kollock and Smith, 2003, Mann and Stewart, 2000). These computer-supported social networks can, indeed, create a sense of community, belonging (Wellman and Gulia, 2003) and can be distinguished by their cultural aspects.

First, CMC was viewed as limited, narrow, depersonalized and self-absorbing (Kiesler et al, 1984, Kiesler et al, 1985; Rice and Love, 1987); aimed to maintain status quo (Dubrovsky et al, 1991) and sometime hostile environment (Sproull and Kiesler, 1986). Others (Jones, 1995, 1999; Kollock and Smith, 2003 and especially Rheingold, 1993) shifted attention from straight comparison of CMC and face-to-face communication to CMC as a special cultural interface with many non-traditional yet socially rich structures. In organizational setting, Walther (1995) showed that CMC users do not experience loss of intimacy, especially not when CMC is synchronous. He also suggested that depersonalization is not necessarily a function of the medium, but relates to the perceived duration of the relationship and the possibility of future communication (Walther, 1994). Hine (2000) contends that when CMC is perceived as a culture and not merely a way of communication, it has provided rich field and easily obtainable data for anthropologists, psychologists, ethnographers alike.

We focus on Internet forums for several reasons. Internet forums are ongoing CMC groups, which in general tend to develop into a specific culture with shared values, accepted behavioural norms and interpersonal relationships (Baym, 1998). Having the option to be able to see (read) communication retrospectively, researchers can map the formation and dynamics of the network from a single, large data set. Many features of social networks (different tie-strengths, personal preferences, likeness, shared interests, interpersonal likes and dislikes) can be qualitatively observed among the regular participants of this virtual community.

The notion of tie strength in social networks

The notion of tie-strength is an important concept in social network analysis. Strength of a tie is a quantifiable property that characterises the link between two nodes. By definition, tie strength is a "combination of the amount of time, the emotional intensity, the intimacy (mutual confiding) and reciprocal services which characterize the tie" (Granovetter, 1973, p1361). The four indicators are thought to be linear combinations of the four elements, positive and symmetric. (Ibid, p.1361). Indicators are actual components of tie-strength (closeness, duration and frequency, breadth of topics and mutual confiding), whereas contextual contingencies (neighbourhood, affiliation, similar socio-economic status, workplace and occupation prestige) are predictors. Predictors are related to tie-strength but not components of it.

Based on Granovetter's weak tie argument (1973), many important claims have been made by a range of people on a range of topics - from job seekers to job providers, social groups to formal organisations, health care systems to drug users, environmental protection to criminology, on topics related to innovation, marketing, migration, maffia and terrorism. Examples of such research can be found in the recent literature (e.g., Albrecht & Hall, 1991; Bian, 1997; Brown & Konrad, 2001; Brown & Reigen, 1987; Burt, 1995; Carpenter et al., 2003; Crowell, 2004; Faia, 2000; Feld, 1997; Greenbaum, 1982; Hagan, 1993; Hansen, 1999; Haines & Henderson, 2002; Jenssen & Koenig, 2002; Karathanos & Pettypool, 1992; Köhler, 2004 ; Krackhardt, 1988; Lavigne, 1996; Lin, 1999; Lin & Dumin, 1986; Macy & Skvoretz, 1998; McGrath et al., 2003; Miller McPherson et al., 1992; Montgomery, 1992, 1994; Morselli, 2003; Rankin, 2003; Schwartz & Sprinzen, 1984; Teorell, 2003; Tindall, 2002; Valente & Vlakov, 2001; Weening, 1993; Weening & Midden, 1991; Wellman & Wortley, 1990; Wilson, 1998; Youm, 2002). However, the proportion of researchers who uses tie-strength is overwhelmingly larger than the number of empirical studies that have made an attempt to measure tie-strength (Mathews et al., 1998). Claims and theories, which rely upon the notion of tie-strength can only be tested if we are able to measure the strength of ties and able to discriminate strong ties from weak ones, independently of the original assertions (Granovetter, 1973).

In the past thirty years of social network analysis, many attempts have been made to find valid indicators and predictors of tie-strength (Walker et al., 1993). The simplest way was to assume that close friends have strong ties and acquaintances or distant friends are connected by weak ties (Erickson et al., 1978; Granovetter, 1974; Murray et al., 1981; Wilson, 1998). Additionally, multiplexity was also used as a strength indicator (Granovetter, 1973). For measuring tie-strength, frequency of contact has been proposed by Granovetter (1974) and Lin et al. (1981), used by Benassi et al. (1999); and reciprocity was suggested by Friedkin (1980). Emotional support offered and received within a tie also proved to be a plausible indicator of tie-strength (Lin et al., 1985; Wellman, 1982; Wellman & Wortley, 1990). Contextual factors such as social homogeneity (Lin et al., 1981), shared affiliation and social circles (Alba & Kadushin, 1976; Beggs & Hurlbert, 1997) were also looked at in reference to tie-strength in social networks.

Marsden and Campbell (1984) investigated two major elements, indicators and predictors of tie-strength. Of all indicators, Marsden and Campbell (1984) showed that many indicators, including frequency and time spent, are contaminated by situational factors (predictors), except one. The measure of closeness was found free of contamination.

In 1998, Mathews and colleagues repeated Marsden and Campbell's study (1984), using a 13-item scale assessing tie-strength with a college student population. Information regarding potential predictors (gender, age, relative, roommate, attending the same lectures, same hometown, overlapping affiliation and duration of relationship) were also collected. The 13 items were related to four factors, namely: intimacy, time, services and intensity. As the aim of Mathews and colleagues (1998) was to find predictors and indicators of tie-strength, they did not make an attempt to quantify strength, nor to make distinction between strong and weak ties. Rather, the relationship between indicators and predictors were investigated and Marsden and Campbell's argument (1984) about contaminated indicators was supported. Evidence was found, however, that certain sets of indicators explain more or less of the variability of the data set (Mathews et al., 1998, p.1463) with intimacy being the strongest indicator of tie-strength.

In addition to the above measures, voluntary investment in the tie, desire for companionship and frequent "meeting" with the tie partner in various context and intimacy can also be used as an indicator of tie-strength (Blumstein and Kollock, 1988; Mitchell, 1987; Plickert et al, 2005), where obviously strong ties exhibit all of the mentioned characteristics, whereas weak ties are mostly lacking these elements. In addition to the indicators discussed, a fairly comprehensive list of potential tie-strength components in various settings is summarised in Table 1.

Table 1. Summary of tie-strength components

Measures	Category	References
Frequency	Indicator	Benassi *et al.*, 1999; Blumstein & Kollock, 1988; Granovetter, 1974; Lin *et al*, 1981; Marsden & Campbell, 1984; Mathews *et al*, 1998; Mitchell, 1987, Perlman & Fehr, 1987
Intimacy/Closeness	Indicator	Blumstein & Kollock, 1988; Marsden & Campbell, 1984; Mathews *et al*, 1998; Mitchell, 1987; Perlman & Fehr, 1987
Voluntary investment in the tie	Indicator	Blumstein & Kollock, 1988; Perlman & Fehr, 1987
Advice given/received	Indicator	Mathews *et al*, 1998
Desire for companionship	Indicator	Blumstein & Kollock, 1988; Perlman & Fehr, 1987
Multiple social context (breadth of topics)	Indicator	Blumstein & Kollock, 1988; Granovetter, 1973; Marsden & Campbell, 1984; Perlman & Fehr, 1987
Long period of time (duration)	Indicator	Blumstein & Kollock, 1988; Granovetter, 1973; Marsden & Campbell, 1984; Perlman & Fehr, 1987
Reciprocity	Indicator	Blumstein & Kollock, 1988; Friedkin, 1980 Granovetter, 1973; Mathews *et al*, 1998; Perlman & Fehr, 1987
Provide support/emotional intensity	Indicator	Blumstein & Kollock, 1988; Granovetter, 1973; Mitchell, 1987; Perlman & Fehr, 1987; Wellman, 1982; Wellman & Wortley, 1990
Mutual confiding (trust)	Indicator	Granovetter, 1973; Marsden & Campbell, 1984; Mathews *et al*, 1998
Sociability/conviviality	Indicator	Mitchell, 1987

Based on the literature we are aware of, quantitative, continuous measure of tie-strength has not been used. Often, researchers use the notion of weak or strong ties (e.g., Feld, 1997; Friedkin, 1980, 1982; Haythornthwaite, 2002; Roch et al., 2000;) as grouping variables. In many papers, it was rather unclear how the researchers obtain information regarding the strength of interpersonal ties. Few notable exemptions are, for instance, Hansen (1999), Harkola and Greve, 1995), Mathews et al. (1998), Plickert et al. (2005), Podolny (2001), van Alstyne and Bulkley (2005) and Wellman and Frank (2001).

Even in research projects, where the authors quantified their tie-strength related variables in their data set (e.g., Mitchell, 1987; Plickert et al, 2005; Wellman & Frank, 2001), the final outcome, again, was nominal data, unsuitable for many statistical analysis, including sophisticated graph theoretical methods available for weighted graphs. Also, there are measures of tie strength which apply in case of particular networks, e.g. economic networks (Podolny, 2001), yet they do not correspond to the strength of social bonds, rather to economic interests, thus can not be applied outside their original context.

Tie-strength measures in virtual communities

Indicators and predictors summarised in Table 1 have been extracted from data collected in off-line social groups and as such, they may or may not be valid in virtual communities. Virtual communities are created/maintained and held together by computer-mediated communication (CMC), therefore components such as help provided and received, time spent together or even communication may have different meanings.

Studies focusing on tie-strength in true virtual communities are rather sparse. Among the few, Muncer et al. (2000a, 2000b) simply defined tie as having at least one posting between two participants and used the number of postings on each strand and frequency to indicate strength. Paolillo (2001) analysed the context of the messages and used informal 'speech' (i.e., using 'u' when writing 'you') and spelling as indicator of friendship and closeness. Adamic & Adar (2003) tested similarities, homepage links and email distribution lists' membership to predict relationships and found that homepage links and mailing lists (except religious lists) are poor predictors of a relationship between two people, whilst having mutual friends seemed to foster relationship developments. Therefore, the aim of this study is to test the above predictors and perhaps identify new ones for tie strength measures that recognise the uniqueness of virtual communities.

METHODS

In case of a systematic network analysis (such as in this project), preferably the entire network is surveyed. Network data relating to tie-strength was collected by survey methods with questionnaires, followed up by virtual 'focus group' discussion to verify the questionnaire results. The 12-question survey used nomination technique with non-specific aided recall. Respondents completed the questionnaires on-line.

Potential problems with reliability of network data are due to the problem of recall and informant accuracy. Killworth, Bernard and Sailer in their series of informant accuracy (Bernard et al., 1980, 1982; Killworth & Bernard, 1976, 1979; Bernard & Killworth, 1977) warned researchers to exercise great caution when interpreting survey data. They claim that people's recall of past communication patterns is far from being accurate and currently used techniques cannot help this problem. On the contrary, Freeman et al., (1987), showed that respondents are able to recall and correctly report relations in general and have a fairly accurate picture about the social relations surrounding them (Freeman et al., 1989, Freeman, 1992). The questionnaire was designed to ask general information, rather than specific actions or occurrences. Because respondents were asked to name people with whom they share friendship, good times, help, trust and confiding, there is another problem we had to face: forgetting in recall-based elicitation (Brewer, 2000). To counterbalance this problem, the full list of topic participants' nicknames was provided. Nicknames were listed in alphabetical order. Answers were recorded by clicks on nicknames.

At the beginning, invitation was sent to everyone who ever posted message on the forum, using the private message function. A reminder message was sent to those who had not completed the questionnaire after the first three weeks. Technically, the system allowed new participants' names to be added to the list as they completed the questionnaire, permitting newcomers to join the sample even after the data collection had started. In reality, there was only a few that joined the sample after the questionnaire was set and due to their relatively little involvement in the group, they had not have formed significant relationships with anyone during the time when the data was collected. The questionnaire was only available for 8 weeks. As it can be seen from Fig. 1, depicting the ratio of active versus all participants, the structure of the on-line community was stable during the data acquisition period.

Validity of the scale was established through methodological triangulation (Creswell & Miller, 2000; Denzin & Lincoln, 1998; Patton, 2002; Strauss & Corbin, 1990). Specifically, between-method triangulation was chosen based on the assumption that each method used in this research will complement, rather than compound, the other methods' strengths and weaknesses (Jick, 1979). Therefore, quantitative tie-strength data from the questionnaire was first checked against information resulted from the prolonged observation (qualitative data) and then discussed with the participants themselves. The latter one proved to be very useful for both parties and provided evidence of the validity of the quantitative data obtained via the questionnaire.

If tie-strength can be objectively quantified, any attempt to measure it should yield various strengths of ties. Correspondingly, any given person's ties to different people may vary in strength and in case of reciprocal ties; the tie-strength does not need to be equal. In order to capture the essence of social networks, classifying the ties according to their strength should allow more variability than the simple dichotomy of strength and weakness. We computed the tie-strength measuring scores as follows: awarded 1 "point" for each nomination (that is when a person received a nomination from any member of the community) and double weighted if the nomination was mutual (pairwise) to recognise reciprocity. Double weighing meant to take into account that a tie from A to B should be stronger if the same tie is confirmed by B. Therefore, 'stars' (actors with many nominations) can have strong ties only if the tie represents mutual acknowledgement. In other words, stars must nominate back those who originally gave them incoming nominations to obtain high scores. Furthermore, the literature suggests that reciprocity is one of distinctive features of strong ties (Granovetter, 1973).

Summated scores were subject to both factor analysis and hierarchical clustering. The results of factor and cluster analysis from the development sample were tested under two condi-

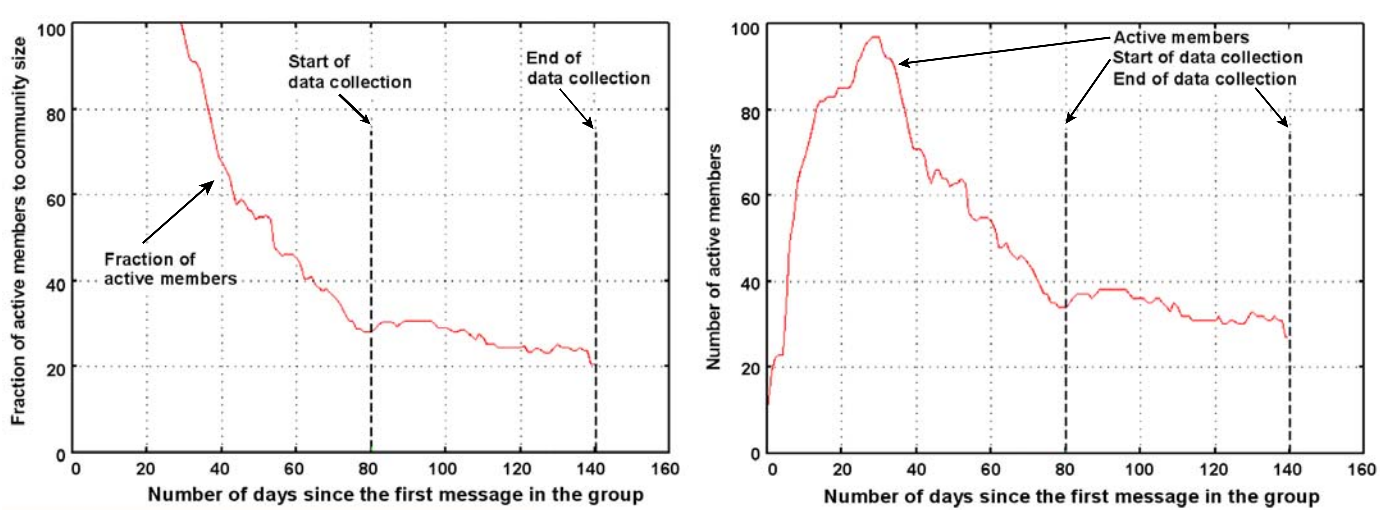


Figure 1. Stability of the discussion forum used for scale-development. On the left panel Absolute number of active participants (defined as those who sent at least one message in 30 days), left panel. Relative number of active participants (defined as the absolute number of active participants/number of all participants up to that date.), right panel. The period of data acquisition is shown with vertical lines.

tions: using data from asymmetric relationships and data with symmetric relationships only. Scale reliability was established on both data sets by Cronbach α values. Factor analysis and hierarchical clustering were performed, and scale reliability coefficient was calculated using SPSS 12.1. We also used Mathematica 5 and Perl scripts to compute values of the tested predictors.

DATA COLLECTION

Although there are many ways to define virtual community (e.g., Hine, 2000; Jones, 1995, 1999; Paccagnella, 1997; Preece, 2000; Rheingold, 1993), for the purpose of this research, virtual communities were defined by using Jones's (1997) four criteria: minimum level of interaction (1) among the variety of communicators (2) via a common public place (3) where most of the CMC occurs and a minimum level of regular membership (4). The selected two groups conform to the above criteria.

Data was collected via an on-line questionnaire in which participants were asked about their relationships to other members in the community. Following the recommendation of Ferligoj and Hlebec (1999), tie-strength was measured with several questions tapping into two distinct levels of relationship (acquaintances and friendship, see Table 4) and summated scores were calculated. The importance of reciprocity was recognised by adjusted weighting of the score in which mutual recognition occurred. The notion of weighting ties has been already used in Podolny (2001) in relation to inter-organisational economic ties, but to our knowledge, not in personal social networks.

To begin, a single topic was selected with 14,907 postings at the time when our data collection ended. The topic has grown out from a discussion on similar childhood memories and later turned into a virtual meeting place for people (mostly middle-aged women with family responsibilities and full time jobs). One of the researchers has been a member of the selected topic since the beginning. The researcher's involvement can be described as complete participation (Spradley, 1980). The researcher virtually 'lived' among the group members for over six months: made appearance every day on the forum, listened to their stories and shared own stories with them, provided encouragement, advice or help as needed and gratefully accepted the reciprocated favours, whilst every effort was made to remain an active yet objective and observant participant. Following the recommendations of Sharf (1999) regarding participant observation on the Internet, the investigator's true identity was revealed prior to data collection (i.e., after the first 3 months period). The decision about when to start collecting data was based on participant observation. Network structure stabilised after the first 1500 – 2000 postings, or approximately 3 weeks (Nepusz et al., 2005).

The personal, prolonged involvement of the researcher was twofold. First, the aim was to develop a sense for the virtual life and particularly for this virtual community. The secondary intention was to establish rapport with and gain trust among the discussion forum members. Fifty-six members of the active 83 participated in this study. As Jones (1999) pointed out, it is seductively tempting to plunge into the vast amount and easily accessible data internet chat forums and discussion boards provide and harvest it for research purpose – without a real understanding of its meaning. In fact, understanding virtual communities is not any easier than to understand face-to-face human interactions. Living among those we wish to understand is a proven method in ethnography and virtual ethnography should not be any different in this respect.

The second set of data was collected from the same web-portal asking volunteers from another long-lived topic to complete the questionnaire. There was no overlap between the two sample groups. Although the topic was smaller (postings at the time of data collection remained under 3000), it had many regular members visiting the topic often to ask and provide help and encouragement relating to their post graduate studies. Members of this topic are all mature students (females, often with full time jobs and/or family responsibilities) of the same distance-learning programme. Sixteen of the total 86 agreed to participate in our study. (The lower positive response rate shows that establishing rapport among participants is just as essential in virtual groups as in their face-to-face counterparts.)

Table 2. The questionnaire (questions as they appear here are translated from Hungarian)

	Question	Measure
1	Which participants of the forum do you like?	Positive relationship
2	Which participants of the forum you do not like?	Negative relationship
3	Which participants do you trust (for example they know your real name, email address, password to your introduction sheet)?	Trust
4	Which participants have trusted you (have seen their introduction sheet, known their real name and email address)?	Trust
5	Which are the forum participants who have asked your help or asked a favour?	Support
6	Which are the forum participants, from whom you asked a favour or you asked their help?	Support
7	Who are the forum participants from which you feel you could ask a favour?	Support
8	Who are the forum participants with whom you have private correspondence?	Intimacy
9	Which of the forum participants do you consider to be your virtual friend?	Companionship
10	Who are the forum participants, with which you discussed topics other than the forum's topic?	Multiplexity
11	With whom of the forum participants, would you like to have a discussion about topics other than the forum's topic?	Multiplexity
12	Which ones of the forum participants would you like to meet in person?	Companionship/ closeness

STATISTICAL ANALYSIS OF THE QUESTIONNAIRE DATA

Although Granovetter (1973) assumed that all ties are symmetric, our data suggest otherwise (see Table 3). In asymmetric relations, strength of tie A → B is not equal to strength of tie B → A. In other words, an asymmetric relationship acknowledges that a relationship between two nodes is not necessarily mutual. Based on Granovetter's work (1973) it can be assumed that strong relationships should contain mutual elements, whilst weak ties can be asymmetric relationships. If we adhere to mutuality of relationships where every indicator of tie-strength must be reciprocal to be acknowledged, tie-strengths are organised in symmetric matrices, where strength of A → B is the same as strength of B → A, thus it can be noted as strength of A → B. The measurement tool (questionnaire) was examined under both conditions.

We considered a pair to be symmetric when mutual nomination was received on the same question(s). For the purpose of comparison we display the symmetric and asymmetric pairs (Table 3). One should note the difference in symmetric and asymmetric nominations. As Table 3 shows, some aspects of a tie are more sensitive to symmetry constraint than others. Notably question 2, which is the only negative question (dislike) does not have symmetry at all. Questions related to trust (Q3 and 4) and expressed feeling of friendship (Q9) seem to be prone to symmetry, whereas questions regarding individual likes, interest and desire (Q1, 10, 11 and 12) are greatly unidirectional, which results significantly larger number of asymmetric pairs than symmetric pairs. The discrepancy between the total number of ties and the sum of ties is due to the large number of non-nominations.

Principal component analysis was performed to identify common factors among the 11 questions. Q2 was not used in the final analysis, as it was the only negative question and as such, showed distinct difference from the acquaintance and friendship factors. The Kaiser-Meyer-Olkin sampling adequacy index was excellent (0.939) when using asymmetric matrix and very good (0.822) when using the more restrictive, symmetric dataset. Factor loadings are shown in Table 4. Factors with eigenvalues greater than 1 were considered. To allow components to be related, Promax rotation was used. Factor analysis identified two factors that were, indeed, significantly correlated (Spearman's r = 0.5). The first factor relates to issues of "liking" someone: shared interest, degree of intimacy and trust, desire to meet in person. The second factor consists of serious "friendship" questions: acknowledgement of the friendship and reciprocal help. Combining the two components of a tie (acquaintances and friendship), scores received on each element were added together. Thus the maximum score for the combined tie-strength was 22, (11 questions, all double weighted), the minimum was zero.

Table 3. Frequency of nominations obtained in the survey

Question	Frequencies: Symmetric 175 pairs	Frequencies: Asymmetric 3080 pairs
1	137	230
2	0	26
3	61	79
4	65	76
5	15	65
6	12	37
7	51	100
8	56	90
9	30	54
10	70	164
11	65	268
12	69	199

The 11 questions were subject to cluster analysis. Hierarchical clustering using squared Euclidean distance and Ward methods produced two clusters that are almost perfectly aligned with the two factors (for comparison, see Table 4). In our case hierarchical clustering separates the questions in the positive quadrant of a two dimensional space. The coordinate axes are labelled with factor names. Graphical representation of the cluster formation is displayed in Fig. 2 and Fig. 3.

Table 4. Summary of the factor loadings under the two conditions.

Questions	Asymmetric (N_{pairs} = 3080) Factor loadings [a] PC 1	PC 2	Cluster	Symmetric (N_{pairs} = 175) Factor loadings [a] PC 1	PC 2	Cluster
1	**.801** [b]	.444	1	.186	.183	1
2	-	-	-	-	-	-
3	**.837**	.698	1	**.790**	.450	1
4	**.849**	.703	1	**.802**	.477	1
5	.523	**.844**	2	.262	**.832**	2
6	.463	**.866**	2	.319	**.799**	2
7	**.824**	.591	1	**.757**	.329	1
8	**.817**	.654	1	**.686**	.498	1
9	.724	**.790**	2	.666	**.720**	2
10	**.749**	.548	1	.306	**.569**	1
11	**.766**	.384	1	**.574**	.167	1
12	**.831**	.483	1	**.633**	.164	1
Cronbach α	*0.9209*	*.7864*		*.7645*	*.7662*	

[a] Promax rotation [b] Numbers in bold indicate factor loadings on the first factor

Although the VTS-Scale was investigated under both conditions (symmetric and asymmetric pairs), the great discrepancy between the number of ties must be noted. Granovetter (1973) and many who followed his footsteps assumed the pairs to be symmetric. The assumption may hold to a certain degree when ties only take dichotomous values: exist or does not exist; positive or negative; and perhaps weak or strong. When one attempts to measure the strength of the tie and represents this strength by a value (that is, the measurement is taken on at least an interval scale), it is unlikely that two people feelings toward each other are exactly the same. Realistically, such scenario only happens on two extreme ends of the scale: on the lowest end, where the value of tie strength is zero (two people do not have any relationship) and maybe on the highest end, where two people benefits from very strong bond between them. In our data, there were 2323 pairs with tie-strength value of zero (75.4% of all pairs) and only 4 pairs (0.1%) with the maximum score of 22. Distribution of the tie- strength values is shown in Fig. 4.

As it can be seen from the dendrogam (Fig. 2 and Fig. 3), questions 5, 6 and 9 are clearly separated from the remaining questions. This small cluster is identical to the component named friendship, whereas the remaining questions form the other, acquaintances component. The results suggest that tie strength may consists of distinguishable elements, thus the VTS-Scale may further be divided into subscales.

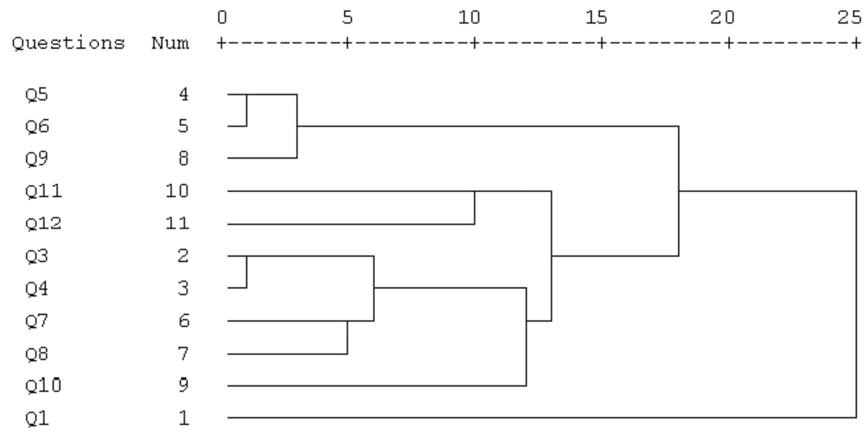


Figure 2 Result hierarchical clustering of VTS-Scale questions (symmetric pairs)

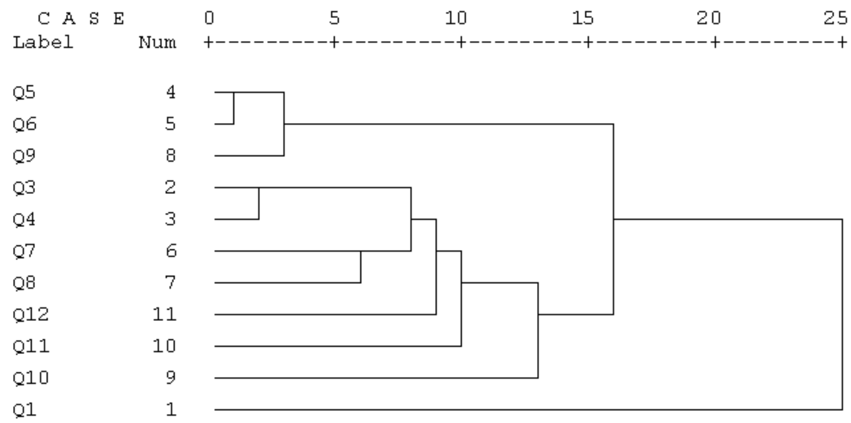


Figure 3 Result hierarchical clustering of VTS-Scale questions (symmetric pairs)

DISCUSSION

Based on the 56 participants who completed the survey, we created a table consisting of 3080 edges possible in principle among the respondents. The ties were grouped as follows: scores in the range 0-3 meant no or weak tie (2756 ties, 89.5%), ties with scores in the range 4-16 were named as medium strong (274, 8.9%), and ties with scores 17-22 were considered strong (50, 1.6%). These ratios seem to contradict the assumption that in virtual communities, the number of strong ties should be higher because it is easier to develop and manage strong ties in situations where face-to-face interactions is not required whilst asynchronous interaction and giving parallel attention is possible.

The cut-off point of 17 was selected because score 17 is above the upper quartile. Thus, one may obtain at most 16 points for the recognitions of acquaintance and 6 points for friendship. In reality, it seems that there is no precise distinction between medium weak and medium strong relationships. Although there is no doubt that such relationships exist, we assume that either they are transitional and usually indicate a (possibly unsuccessful) attempt to form a strong relationship, or they are the indicators of a one-sided relationship (cases when someone nominating a 'star' of the community). Hence we did not subdivide the medium strong group further. The dynamics of these medium strong ties is, indeed, very interesting and probably imperative in network formation; therefore the issue will be further investigated in future research.

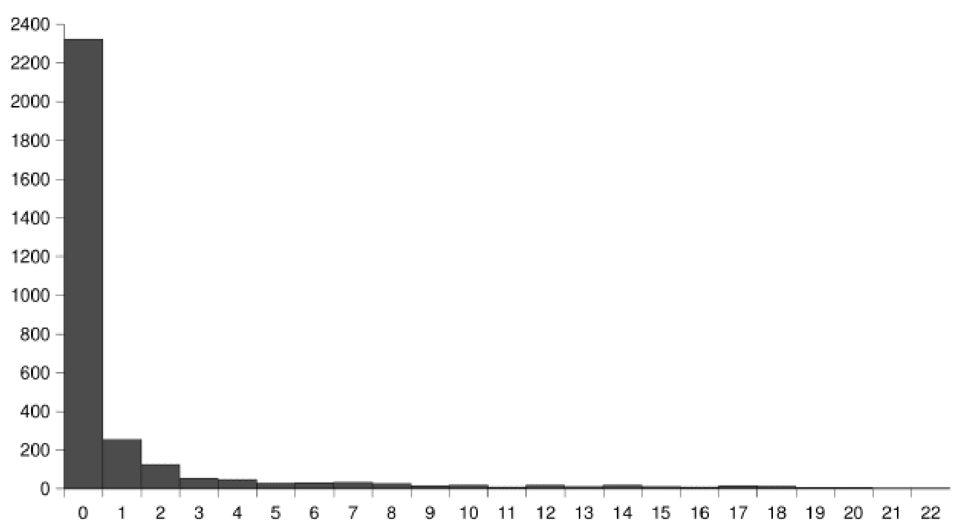


Figure 4 Distribution of tie-strength values in the asymmetric pairs

Granovetter (1973) assumed that ties between friends and acquaintances are likely to differ in strength. In short, friends are connected with strong ties, whereas acquaintances are connected by weak ties. Correspondingly, in our data, tie-strength appears to be comprised of two components: degrees of acquaintance (interests, liking, private communication) ranging from knowing each other to casual friendship and close friendship with mutual acknowledgement of the position, multiplexity and reciprocal support (help, advice, assistance, emotional support). This result corroborates Tausing and Michello's (1988) findings that people prefer to seek support via their strong ties, regardless of the nature of the problem.

From the measurement viewpoint, every strong tie should include a weak tie as well (i.e., friendship builds upon acquaintances). In other words, it is not possible for someone to have a strong tie (reciprocal confiding, mutual acknowledgement as friends) without claiming a weak tie with that particular person. Yet, not all relationships reach the level of friendship – some may stay as acquaintances over time, whereas others may fade out eventually. Conditional probabilities (Fig. 5) suggest that acquaintance is part of friendship. It can be read from Fig. 5 that the probability to have at least 13 points from a maximum of 16 for those who obtained the maximum score of the friendship (6) was 66.67%. On the contrary, the probability of having the maximum 6 points for those who scored 13 on the acquaintance-scale is only 24.24%. Thus we can conclude that strong friendship must contain the elements of acquaintance, but in most cases strong acquaintance does not mean friendship.

The condition of a strong relationship was having at least 1 point from the friendship component (Q5, 6 and 9). Thus, strong relationship is defined as points between 17 and 22. As Fig. 6 shows, increasing strength of friendship not only results in increased statistical means of acquaintances, but also produces smaller range and reduced variability.

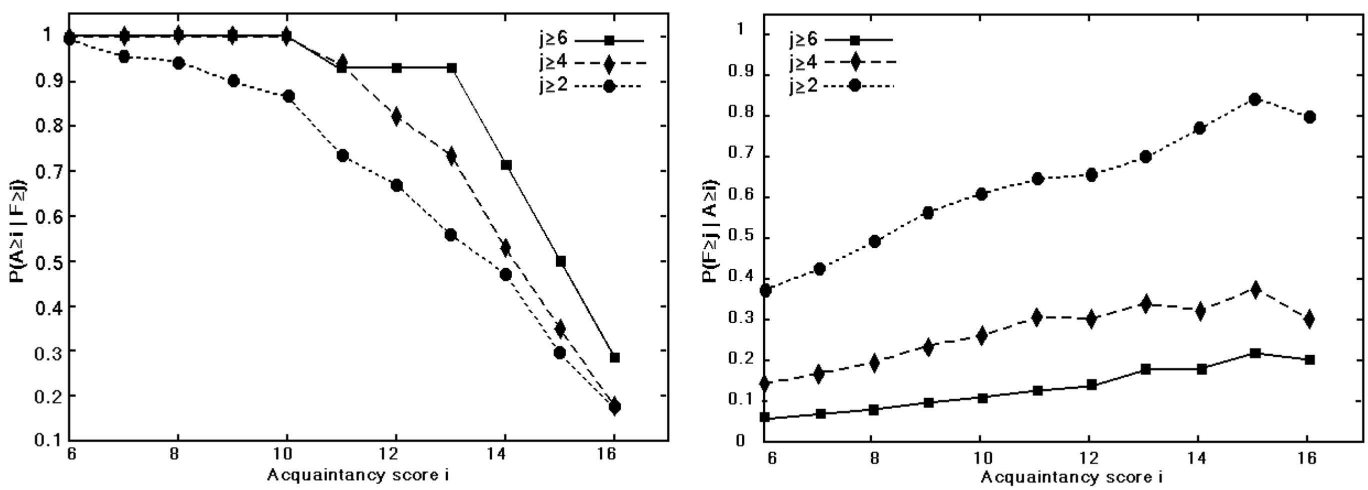


Figure 5. Conditional probability of acquaintance score given the friendship score, left panel, conditional probability of friendship score given the acquaintance score right panel.

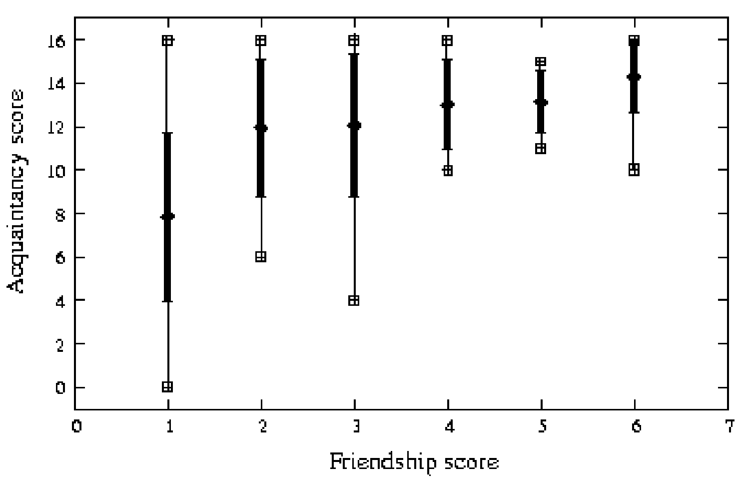


Figure 6. Summary of acquaintance scores given the friendship score. For a given friendship score displayed are: the mean, standard deviation around the mean, minimal and maximal acquaintance scores.

Two thirds (20) of all the strong ties were reciprocal (i.e. reciprocal tie with the same strength), 10 were non-reciprocal; whilst for the medium strong ties we found 96 non-reciprocated ties and 89 reciprocal connections. Judging from the content of Table 3 (question 9), developing friendship tie may not be any easier in virtual life as claimed by, for instance, Haythornthwaite et al. (1995), Kollock and Smith (2003) or Wellman and Gulia (2003). Even if developing and maintaining friendship ties via CMC is easier, it does not necessarily mean that people will have more ties when they are on-line, let alone stronger on-line than off-line ties. This finding supports Holme et al.'s (2004) recent discoveries regarding characteristics of on-line communities. Between all possible pairs of the 56 respondents in our data set, only 1.6% had strong friendship ties and an additional 8.9% developed into acquaintances. The remaining 89.5% of all ties were practically non-existent. However, this is in concordance with our preliminary expectations. The total number of possible pairs is just a theoretical limit and one cannot reasonably expect that all participants in a community form ties with each other. This shows that virtual communities, in this sense, are similar to their off-line counterparts. As we consider only a chosen few people as our friends or acquaintances from the vast amount we meet each day on the streets or in clubs and parties, we also carefully choose from the ones we cross path in cyberspace.

In terms of numbers of strong, medium-strong or weak ties a person has in this particular community, we found that the average numbers of strong, medium and weak ties were 1.75 (±3.15), 9.66 (±9.53) and 103.65 (±5.83), respectively. The total number of all possible ties related to a given participant is 110 (sum of 55 outgoing and 55 incoming edges). The adjacency matrix of the forum participants was sparse: of all the 17030 possible edges 1147 were present (6.74%). The adjacency matrix of the subnetwork comprised of those forum participants was also sparse, out of all 3080 possible edges 582 were present (18.90%).

Marsden and Campbell (1984) found closeness to be the best indicator of tie-strength (time spent together and duration being good but contaminated indicators), whereas in Mathews et al. (1998), intimacy appeared to be the most important factor explaining variability in tie-strength, followed by time, reciprocal services and intensity. In our survey, questions related to explicitly stated friendship and mutual help seemed to make the distinction between weak and strong ties. Although the three measures cannot be directly compared, it can be said that if mutual acknowledgment of the friendship is measured as the ratio of reciprocated ties, it is most pronounced in the strongest tie group.

VALIDITY AND RELIABILITY

Values of the tie-strength measures using the developmental sample (N = 56) were used to create graphs showing (Fig. 7) people's relationships within the group. Discrepancy in numbers is due to the way data was collected. As we used aided recall, respondents were provided with the complete list of forum participants (116), not only those who completed the questionnaire.

Community structure was explored using the Markov Clustering method (van Dongen, 2000a,b). This method discovers clusters by simulating a large amount of random walks on the directed edges of the graph. The main concept is based on the observation that random walks initiated from densely connected clusters tend to remain in the same cluster. We used the original implementation of the author, available as a separate software package under Debian Linux (http://micans.org/ mcl/).

Freeman (1992) and his colleagues (1989) showed that people have a fairly accurate picture of their immediate social world and their mental images about their group's structure are closely corresponding with observed interactions. The mental process of creating a social map of this particular on-line group was further assisted by the fact that – due to the nature of the discussion forums – large part of the interaction among group members is visible to all. Therefore it is easier for individuals to correctly judge other people's relationships. Discussion with group members supported the results presented in a graph thus provided evidence for validity of our measures. In general, participants agreed with their network position, yet in some cases their perception about the amount of their strong ties was slightly overestimated. As stated in the method section, one of the authors was an active member of this on-line group (number 8 on Fig. 7). As the third point of the between-method triangulation, results were checked against the active participant researchers' field notes and personal perceptions of the group structure as well. In summary, network structure displayed in Fig. 7 'made sense', meaning what is on the graph is congruent with the researcher's personal impression of the group as a whole and it's members' place within it.

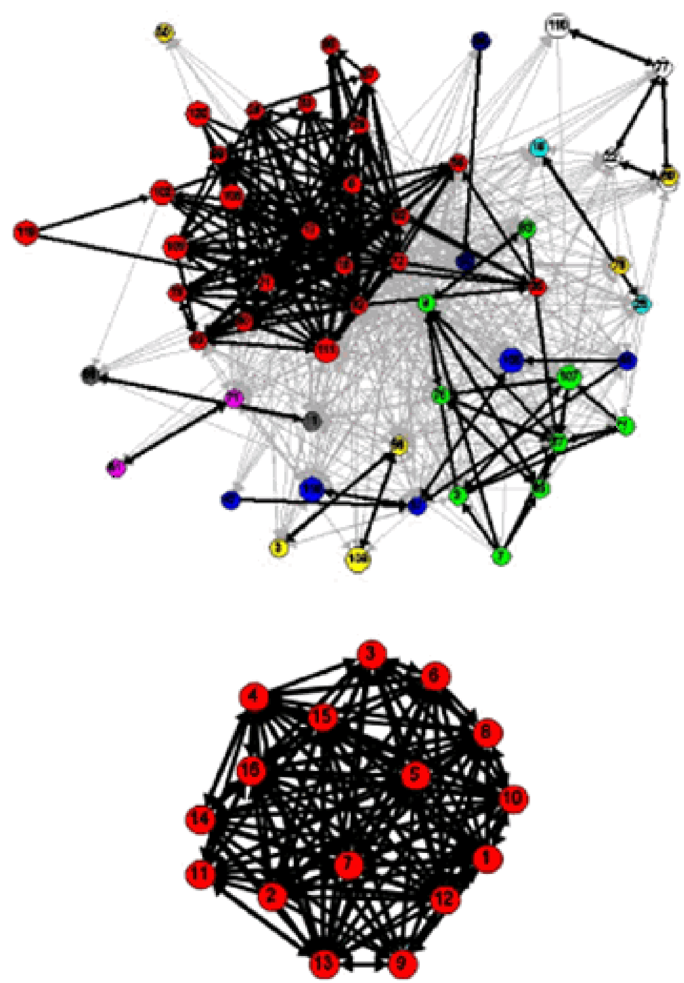


Figure 7. Top: Network structure revealed by Markov clustering based on tie-strength measures (Development sample, Npairs = 3080). Clusters are coded with colours. Bottom: Network structure based on tie-strength measures (Test sample, Npairs = 240).

In the network shown in Fig 7., on the left panel, black, bold lines represent ties within a cluster, whereas grey lines shows relationship between clusters. Nodes with the same colours belong to the same cluster. The nodes coloured yellow are one-member clusters. The 'hard core' of this particular on-line group is the one in red. This is not only the largest cluster within the network, but also the most strongly connected. They meet and communicate regularly over the internet, some of them also met in person. Isolated pairs with strong tie between them are typically people know each other from another topic. They tend not to connect strongly to the main group.

VTS-Scale's reliability first was estimated for the developmental sample (N = 56) by calculating reliability coefficients (Table 4). Cronbach α s using asymmetric and symmetric matrices for the 11 questions together were 0.92 and 0.81, respectively. Reliability coefficients were also calculated for the test sample (N = 16), where Cronbach α for the 11-question scale had an excellent value of 0.86.

Subscale reliability was reassuring for the acquaintance-factor (α = 0.85) but alarmingly low (α = 0.52) for the friendship network. One plausible explanation for the low value may be found in the content (Question 5, 6 and 9). Question 5 and 6 are related to asking and providing help, which are indicators of a close relationship in a friendship network (developmental sample) but does not quite carry the same weight in a group, where the main reason of the group's existence is to provide help to each other (test sample). For the same reason, the relationships in the second group tend to be more functional. Personal feelings (likes and dislikes) are less relevant. Yet, people in the later group indicated in their answers that in general, they do like each other. The relative number of nominations for the first question was much higher than the same in the first sample.

As expected, the test sample shows great cohesion among members, Fig. 7, right panel. All 16 members form one cluster. Fig. 8 shows that the distribution of tie-strength scores is somewhat different and the mean score is higher for the test sample than for the development sample. This is due to the fact, that the test group was smaller, and respondents happened to know each other much better than in the original group, as reflected in the relative number if non nominations (12% for the test sample and 75% for development sample). Our findings (cf. Fig. 6, right) are in concordance with our view about the mutual supportive nature of the second group. The VTS-Scale seemed to be reliable and valid measurement tool for tie strength in virtual communities. Due to the obvious difference between the two groups, component of tie-strength require further investigation. It is recommended to use the VTS-Scale as a one-dimensional measure and sum scores from Q1 to Q12 (not using Q2). Factor loadings of a one factor solution is better than acceptable (Table 5), ranging between the good 0.52 and excellent 0.87. High Cronbach α values (> 0.85) provided further reassurance about the VTS-Scale.

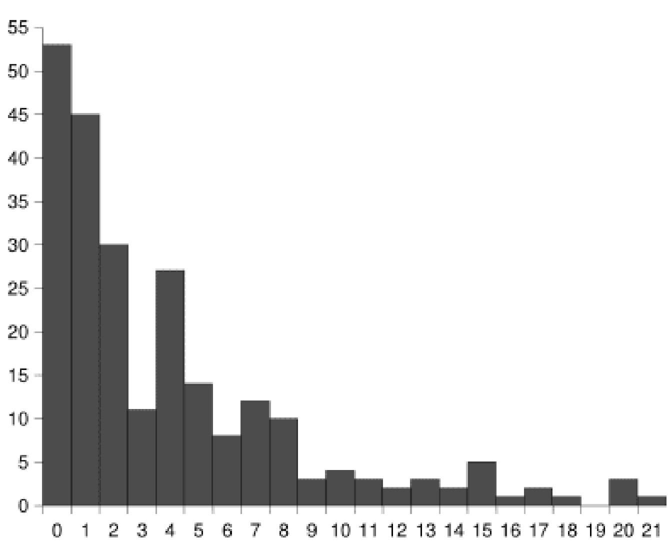


Figure 8. Distribution of tie-strength values in the test sample.

Table 5. Summary of the factor loadings in two independent samples.

Questions	Factor loadings	
	N_{pairs} = 3080	N_{pairs} = 240
1	.751	.630
2	-	-
3	.859	.842
4	.870	.775
5	.668	.574
6	.629	.556
7	.815	.522
8	.830	.827
9	.803	.675
10	.756	.655
11	.706	.573
12	.786	.698
Cronbach α	*0.924*	*0.859*

CONCLUSION

Based on our result, we can safely conclude that indicators in virtual groups are similar to those in off-line networks. Trust, mutual confiding, multiplexity and shared interests is equally important in both types of social groups. The unique aspect of virtual communities is related to help asked and provided; and the desire to meet in person.

Help is easily available on the net. Posting a message on a list serve, an electronic board or initiating a new topic in a discussion forum environment pleading for help is sufficient to solicit advice, help or even emotional support. A person who is in need of advice can rely upon previous personal experiences that somebody will be out there to respond. As no relationship is required between the person who seeks help and a person who is willing to provide it, such network formation can only be described as an association network. Reciprocity has a different meaning in such networks as help or advice may very well be provided by a third (fourth, fifth, etc.) person or group within a broad timeframe. It is not expected to help the particular person who helped us when we needed, but to provide help or support to those who needs it within that particular virtual culture

(Wellman and Gulia, 2003). Therefore, receiving virtual support is not the equivalent to receiving personal support or help through a friendship tie. As a beggar on the street cannot claim to have a large friendship network based on the fact that a hundred people donated their change, support offered to complete strangers cannot be compared to support asked for- and received, or offered between two friends. To use help as indicator, one must step beyond what is normally offered on the net: either mutual reciprocity or help asked and provided beyond what can be normally expected on the net is required to make distinction between strong and weak ties based on help.

Developing the virtual friendship into a traditional relationship seems to be an important step. Even the most devoted members of virtual communities protect their personal life. People give strong value on their true identity (not as much to share personal information and photos, but to actually meet face-to-face) and their time. To make a commitment, moreover, to desire a personal meeting indicates that two (or more) people in this situation are probably strongly connected.

In summary, the VTS-Scale appears to be a valid and reliable measurement of tie strength, perhaps both in virtual and off-line groups. However, the difference between the two samples highlighted the importance of contextual contingencies. Differences in situations must be taken into account and reflected in the phrasing of the questions. The VTS-Scale is a useful tool for researchers who wish to use tie strength as dependent, rather than independent variable as it provides continuous measures of tie-strength instead on the previously used binary classification. The use of tie strength as an outcome variable, if measured on nominal or ratio scale, is limited to non-parametric techniques.

In addition, studying details of social structure may benefit from weighting network edges on a continuous scale, as in that case a large apparatus of statistical techniques may be meaningfully applied. In case of simply grouping edges into strong and weak ones much less can be said about the social structure as a whole. Future research should aim to explore the friendship and acquaintances components further, both in on-line and off-line communities.

REFERENCES

Adamic LA & Adar E. 2003. Friends and neighbours on the web. *Social Networks, 25*: 211-230.

Alba RD & Kadusin C. 1976. The intersection of social circles: A new measure of social proximity in networks. *Sociological Methods and Research, 5*: 77-102.

Albrecht TL & Hall B. 1991. Relational and content differences between elites and outsiders in innovation networks. *Human Communication Research, 17*: 535-561.

Baym NK. 1998. The emergence of online community. In: Jones S (Ed.), *Cybersociety 2.0* (pp. 35-68). Thousand Oaks, CA: Sage.

Beggs JJ & Hurlbert JS. 1997. The social context of men's and women's job search ties: membership in voluntary organizations, social resources and job search outcomes. *Sociological Perspectives, 40*: 601-625.

Benassi M, Greve A, & Harkova J. 1999. Looking for a network organization: The case of GESTO. *Journal of Market–Focused Management, 4*: 205-229.

Bernard HR & Killworth PD. 1977. Informant accuracy in social network data II. *Human Communication Research, 4*: 3-18.

Bernard HR, Killworth PD, & Sailer L. 1980. Informant accuracy in social network IV: A comparison of cliques-level structure of behavioural and cognitive network data. *Social Networks,* 2: 191-218.

Bernard HR, Killworth PD & Sailer L. 1982. Informant accuracy in social network V: An experimental attempt to predict actual communication from recall data. *Social Science Research, 11*: 30-66.

Bian Y. 1997. Bringing strong ties back in: Indirect ties, network bridges, and job searchers in China. *American Sociological Review, 62*: 366-385.

Blumstein P & Kollock P. 1988. Personal relationships. *Annual Review of Sociology, 14*: 467-490.

Brewer DD. 2000. Forgetting in the recall-based elicitation of personal and social networks. *Social Networks, 22*: 29-43.

Brown JJ & Reigen PH. 1987. Social ties and word-of mouth referral behaviour. *Journal of Consumer Research, 14*: 350-362.

Brown W & Konrad AM. 2001. Granovetter was right: The importance of weak ties to a contemporary job search. *Group and Organizational Management, 26*: 434-462.

Burt R. 1995. *Structural holes: The social structure of competition*. Boston, MA: Harvard University Press.

Carpenter D, Esterling K, & Lazer D. 2003. The strength of strong ties: A model of contact making in policy network with evidence from U.S. health politics. *Rationality and Society, 15*: 411-440.

Creswell JW & Miller DL. 2000. Determining validity in qualitative inquiry. *Theory into Practice, 39*: 124-131.

Crowell LF. 2004. Weak ties: A mechanism for helping women expand their social network and increase their capital. *Social Science Journal, 411*: 15-28.

Denzin NK and Lincoln YS. 1998. *The landscape of qualitative research: Theories and issues*. Thousand Oaks, CA: Sage

Erickson BH, Nosanchuk TA, Mostacci L, & Ford C. 1978. The flow of crisis information as a probe of work relations. *Canadian Journal of Sociology, 3*: 71-87.

Faia MA. 2000. "Three can keep a secret if two are dead" (Lavine, 1996): Weak ties as infiltration routes. *Quality and Quantity, 34*: 193-216.

Feld SL. 1997. Network analysis using the method of path lists: testing the weak ties hypothesis. *Social Networks, 19*: 97-111.

Ferligoj A & Hlebec V. 1999. Evaluation of social network measurement instruments. *Social Networks, 21*: 111-130.

Friedkin N. 1980. A test of structural features of Granovetter's strength of weak ties theory. *Social Networks, 2*: 411-442.

Friedkin N. 1982. Information flows through strong and weak ties in intraorganizational social networks. *Social Networks, 3*: 273-85.

Freeman LC. 1992. Filling in the blanks: A theory of cognitive categories and the structure of social affiliation. *Social Psychological Quarterly, 55*: 118-127.

Freeman LC, Roomney K & Freeman SC. 1987. Cognitive structure and informant accuracy. *American Anthropologist,* 89: 310-325.

Freeman LC, Freeman SC, & Michaelson AG. 1989. How humans see social groups: A test of the Sailer-Gaulin models. *Journal of Quantitative Anthropology, 1*: 229-238.

Garton L, Haythornthwaite C, & Wellman B. 1997. Studying online social networks [Electronic version]. *Journal of Computer Mediated Communication*, 3. Retrieved January 25, 2005 from http://jcmc.huji.ac.il/vol3/issue1/garton.html

Granovetter MS. 1973. The strength of weak ties. *American Journal of Sociology, 78*: 1360-1380.

Granovetter MS. 1974. *Getting a job: A study of contacts and careers*. Boston, MA: Harvard University Press.

Granovetter MS. 1982. The strength of weak ties. A network theory revisited. In Marsden PV & Lin N (Eds.): *Social structure and network analysis* (pp. 105-130). Beverly Hills, CA: Sage.

Greenbaum SD. 1982. Bridging ties at the neighbourhood level. *Social Networks, 4*: 367-384.

Hagan J. 1993. The social embeddedness in crime and unemployment. *Criminology*, 31: 465-491.

Haines VA & Henderson LJ. 2002. Targeting social support: A network assessment of the convoy model of social support. *Canadian Journal of Aging, 21*: 243-256.

Hansen MT. 1999. The search transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly,* 44: 82-84.

Harkola J & Greve A. 1995. Diffusion of technology: Cohesion or structural equivalence? In: *Academy of Management Best Papers Proceedings* (pp. 422-426). Academy of Management Meeting, Vancouver, Aug. 6-9, 1995

Haythornthwaite C. 2002. Strong, weak, and latent ties and impact of new media. *The Information Society,* 18: 385-401.

Hine C. 2000. *Virtual ethnography*. London: Sage.

Holme P, Edling CR, & Liljeros F. 2004. Structure and time evolution of an Internet dating community. *Social Networks*, 26: 155-174.

Jenssen JI & Koenig HF. 2002. The effect of social networks on resource access and business start-ups. European Planning Studies, 10: 1039.

Jick T. 1979. Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly,* 24: 602-611.

Jones Q. 1997. Virtual-communities, virtual settlements & cyber-archaeology: A theoretical outline. *Journal of Computer-mediated Communication 3*. Retrieved March 4, 2005 from: http://ww w. ascus c. org/ jcmc/ vol3/ issue3/-jones. html

Jones S. 1995. Understanding community in the information age. In Jones S (Ed.), *Cybersociety: Computer-mediated communication and community*. Thousand Oaks, CA: Sage.

Jones S. 1999. *Doing internet research*. Thousand Oaks, CA: Sage.

Karathanos P & Pettypool MD. 1992. A network approach to formulating communication strategy. *Journal of Technical Writing and Communication, 22*:95-103.

Kiesler S, Siegel J, & McGuire TW. 1984. Social psychological aspects of computer mediated communication. *American Psychologist*, 39:1123-1134.

Kiesler S, Zubrow D, Moses AM, & Geller V. 1985. Affect in computer mediated communication: An experiment in synchronous terminal-to-terminal discussion. *Human Computer Interaction*, 1: 77-104.

Killworth PD & Bernard HR. 1976. Informant accuracy in social network data. *Human Organization,* 35: 269-286.

Killworth PD & Bernard HR. 1979. Informant accuracy in social network data III: A comparison of triadic structure in behavioural and cognitive data. *Social Networks,* 2: 10-46.

Kollock P & Smith,MA. 2003. Communities in cyberspace. In Kollock P & Smith MA (Eds.), *Communities in cyberspace* (pp. 3-29). London: Routledge.

Köhler B. 2004. Weak ties, modern epidemics, and world society. *Connections* 26, 9-10. Retrieved September 1, 2005 from http://www.insna.org/Connections-Web/Volume26-1/2.Kohler.pdf

Krackhardt D. 1988. Simmelian ties: support strong and sticky. In Kramer RM and Neale MA (Eds), *Power and influence in organisations* (pp. 21-38). Thousand Oaks, CA: Sage.

Lavigne Y. 1996. *Hell's angels: Three can keep a secret if two are dead*., Secaucus, NJ: Carol Publishing Group

Lin N. 1999. Social network and status attainment. *Annual Review of Sociology*, 25: 467-487.

Lin N & Dumin M. 1986. Access to occupations through social ties. *Social Networks,* 8: 365-385.

Lin N, Vaughn JC, & Ensel WM. 1981. Social resources and occupational status attainment. *Social Forces,* 59:1163-1181.

Lin N, Woelfel MW, & Light SC. 1985. The buffering effect of social support subsequent to an important life event. *Journal of Health and Social Behaviour,* 26: 247-263.

Macy MW & Skvoretz J. 1998. The evolution of trust and cooperation between strangers: A computational model. *American Sociological Review,* 63: 638-660.

Mann C & Stewart F. 2000. *Internet communication and qualitative research*, Sage, London.

Marsden PV & Campbell KE. 1984. Measuring tie-strength. *Social Forces,* 63: 482-501.

Mathews KM, White MC, Soper B, & von Bergen CW. 1998. Association of indicators and predictors of tie-strength. *Psychological Reports,* 83: 1459-1469.

McGrath CA, Vance CM, & Gray ER. 2003. With a little help from their friends: exploring the advice networks of software entrepreneurs. Creativity and Innovation Management, 12: 2-10.

McPherson MJ, Popielarz PA, & Drobnic A. 1992. Social networks and organisational dynamics. *American Sociological Review,* 57: 153-180.

Mitchell JC. 1987. The components of strong ties among homeless women. *Social Networks,* 9: 37-47.

Montgomery JD. 1992. Job search and network composition: Implication of the strength-of-weak-ties hypothesis. *American Sociological Review, 57*:586-596.

Montgomery JD. 1994. Weak ties, employment, and inequality. *The American Journal of Sociology,* 99: 1212-1236.

Morselli C. 2003. Career opportunities and network-based privileges in the Cosa Nostra. Crime, *Law and Social Change,* 39: 383-418.

Muncer S, Burrows R, Pleace N, Loader B, & Nettleton S. 2000a. Births, deaths, sex and marriage… but very few presents? A case study of social support in cyberspace. *Critical Public Health, 101*:1-18.

Muncer S, Loader B, Burrows R, Pleace N, & Nettleton S. 2000b. Form and structure of newsgroups giving social support: A network approach. *Cyberpsychology and Behaviour,* 3: 1017-1029.

Murray SO, Rankin JH, & Magill DW. 1981. Strong ties and job information. *Sociology of Work and Occupational Status,* 8: 119-136.

Nepusz T, Bazsó F, & Petróczi A. 2005. *Discovering Community Structure from Communication Characteristics*. Poster session presented at the. NATO Advanced Study Institute: Dynamics of Complex Interconnected Systems: Networks and Bioprocesses workshop, Geilo, Norway.

Paccagnella L. 1997. Getting the seats of your pants dirty: Strategies for ethnographic research on virtual communities. *Journal of Computer-mediated Communication 3*. Retrieved March 5, 2005 from http://www. nicoladoering.de/Hogrefe/ paccagnella97. htm

Paolillo JC. 2001. Language variation on Internet Relay Chat: A social network approach. *Journal of Sociolinguistics, 5*:180-213.

Patton MQ. 2002. *Qualitative evaluation and research methods* (3rd ed). Thousand Oaks, CA: Sage.

Perlman D & Fehr B. 1987. The development of intimate relationship. In D. Perlman & S. Duck (Eds.), *Intimate relationships*, Newbury Park, CA: Sage

Plickert G, Wellman B, & Côte R. 2005. It's not who you know, it's how you know them: Who exchanges what with whom? In Blokland T and Savage M (Eds), Social Capital on the Ground (forthcoming), Oxford, Blackwell. Retrieved September 3, 2005 from http://www.epas.utoronto.ca/~wellman/ publications/reciprocity05 /reciprocity3-5.pdf

Podolny JM. 2001. Networks as the piper and prisms of the market. *American Journal of Sociology,* 107: 33-60.

Preece J. 2000. *Online Communities*. New York: John Wiley & Sons Inc.

Rankin B. 2003. How low-income woman find jobs and its effects on earning. *Work and Occupation,* 30: 281-301.

Rheingold H. 1993. *The virtual community: Homesteading on the electronic frontier*. Reading, MA: Addison-Wesley.

Rice RE & Love G. 1987. Electronic emotion: Socioemotional content in a computer-mediated network. *Communication Research,* 14: 85-108.

Roch CH, Scholz JT, & McGraw KM. 2000. Social networks and citizen response to legal change. *American Journal of Political Science, 44*:777-791.

Schwartz JE & Sprinzen M. 1984. Structures of connectivity. *Social Networks,* 6: 103-140.

Sharf BF. 1999. Beyond Netikett: The ethics of doing naturalistic discourse research on the internet. In Jones S, *Doing internet research* (pp.243-256), Thousand Oaks, CA: Sage.

Spradley JP. 1980. *Participant observation*., Fort Worth, TX.: Harcourt.

Strauss A & Corbin J. 1990. *Basis of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.

Tausing M & Michello J. 1988. Seeking social support. *Basic and Applied Social Psychology,* 9: 1-12.

Teorell J. 2003. Linking Social Capital to Political Participation: Voluntary Associations and Networks of Recruitment in Sweden. *Scandinavian Political Studies, 26*:49-66.

Tindall DB. 2002. Social networks, identification and participation in an environment movement: Low-medium cost activitism within the British Columbia Wilderness reservation movement. *The Canadian Review of Sociology and Anthropology,* 39: 413-452.

Valente TW & Vlakov D. 2001. Selective risk taking among needle exchange participants: Implications for supplemental interventions. *The American Journal of Public Health,* 91: 406-411.

van Alstyne M & Bulkley N. 2005. E-Mail and output: Communication effects on Productivity. Retrieved Septem-ber 3, 2005 from http://catalyst.gsm.uci.edu/tools/ dl_public.cat?year=2003&file_id=116&type=cal &name =ResultsInformationandProductivity. pdf

van Dongen S. 2000a. *Graph Clustering by Flow Simulation*. Unpublished PhD thesis, University of Utrecht, May 2000.

van Dongen S. 2000b. A stochastic uncoupling process for graphs. Technical Report, INS-R0011. National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam, May 2000.

Walker ME, Wasserman S, & Wellman B. 1993. Statistical model for Social Support Networks. *Sociological Methods and Research,* 221: 71-98.

Walther JB. 1994. Anticipated ongoing interaction versus channel effects on relational communication in computer-mediated interaction. *Human Communication Research,* 20: 473-501.

Walther JB. 1995. Relational aspects of computer-mediated communication: Experimental observation over time. *Organization Science*, 6: 186-203.

Weening MW & Midden CJ. 1991. Communication network influences on information diffusion and persuasion. *Journal of Personality & Social Psychology*, 61:734-742.

Weening MWH. 1993. The strength of weak and strong communication ties in community information program. *Journal of Applied Social Psychology,* 23: 1712-1731.

Wellman B & Frank K. 2001. Network Capital in a multi-level world: Getting support in personal communities. In Lin N, Cook K & Burt R (Eds) *Social capital: Theory and research* (pp. 233-273). Hawthorne, NY: Aldine DeGruyter.

Wellman B. 1982. Studying personal communities. In Marsden PV and Lin N, (Eds.): *Social structure and network analysis* (pp. 61-80), Beverly Hills, CA: Sage

Wellman B. 2001. Physical Place and cyberplace: The rise of personalised networking. *International Journal of Urban and Regional Research*, 25:221-252.

Wellman B & Gulia M. 2003. Virtual communities as communities. In Kollock P and Smith MA. (Eds.) *Communities in cyberspace* (pp. 167-195), London: Routledge.

Wellman B & Wortley S. 1990. Different strokes from different folks: community ties and social support. *The American Journal of Sociology,* 96: 558-589.

Wilson TD. 1998. Weak ties, strong ties: network principles in Mexican migration. *Human Organization, 57*:394-405.

Youm Y. 2002. Social network effects on the transmission of sexually transmitted diseases. *Sexually Transmitted Diseases,* 29: 689-697.

CONNECTIONS 27(2): 53-58

http://www.insna.org/Connections-Web/Volume27-2/Portnova.pdf

Another Hundred Days: Social Contacts in a Senior Class[1]

Aleksandra Portnova
Tykhe Capital, LLC, NY

Patti Frazer Lock
Mathematics, St. Lawrence University Canton, NY

Brian C. Ladd
SUNY Potsdam, NY

Christine Zimmerman
Director of Institutional Research, St. Lawrence University Canton, NY

Acquaintance data was collected from the Class of 2002 at St. Lawrence University, using a web-based survey. We discuss the graph statistics of the social network resulting from that data. We also use demographic data tracked on a per student basis to examine how acquaintanceship circles differ between different groups. We look explicitly at males/females, students of color, varsity athletes, the effect of different academic majors, transfer students, and members of fraternities/sororities.

INTRODUCTION

It has been mentioned that the field of social networks is populated by "too many methods chasing too few data". (Martin, et.al., 2001) Recent research at St. Lawrence University attempts to address this problem by collecting data on the social interactions within the senior class at a small liberal arts college. This paper reports the inception of what we hope will be an ongoing project, our first data collection, and some initial findings in that data.

St. Lawrence University, located in Canton, New York, is the oldest continuously operating coeducational institution of higher learning in New York State. It has approximately 2100 undergraduates and a small graduate program in Education (St. Lawrence University, 2003). It is a private, non-denominational, residential, liberal arts college with a strong commitment to both the academic and extracurricular lives of its students.

St. Lawrence University places great importance on the quality of student life and actively seeks input from the students themselves through interaction with student organizations and through student surveys, including an exit survey completed by all graduating seniors. In this atmosphere of information gathering, it made sense to consider capturing information on the social contacts within the senior class.

With funding from a student research grant, one of the authors began a mathematics Honors project that included collecting acquaintanceship relations within the senior class, translating that relationship data to a network graph, and analyzing the characteristics of the network graph. Members of the administration at St. Lawrence University then asked her to expand her work by augmenting the nodes in the graph with internal demographic data and analyzing the interaction between membership in a defined demographic group and relationships with the senior class as a whole.

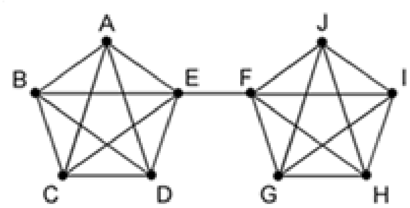
[1] Corresponding author: Brian C. Ladd, Department of Mathematics, Computer Science and Statistics, St. Lawrence University, Canton, NY 13617, phone: 315-229-5293, fax: 315-229-7413, blad@it.stlawu.edu

This paper discusses related work in social networks and graph theory, gives a description of the methodology used to collect the data, and presents the initial results found in analyzing the data.

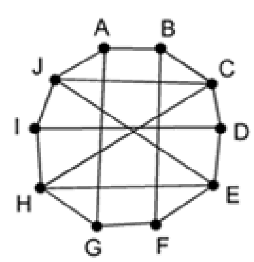
RELATED WORK

The small worlds problem remains an interesting and accessible introduction to complex graph theory. Starting with the simple description of the six degrees of separation, complicated graph theory concepts such as characteristic path length and the clustering coefficient can be introduced. This research focused on collection of data for populating a social network graph as well as on the characteristics of small-world networks.

Our interest in the social networks within our student body grew out of descriptions of the characteristics of such graphs. The small-worlds phenomenon has entered into common knowledge through John Guare's *Six Degrees of Separation* (1990) and the "Kevin Bacon Game" (Reynolds, 1999) complete with a Visa television commercial. As a New York Times article recently noted "Network theory is hot." (2003). Thus, it was relatively easy to explain this research to the administration of St. Lawrence University and pique their interest as well.



Number of vertices = 10
Radius = 2, Diameter = 3
Characteristic path length = 2
Clustering coefficient = 0.92



Number of vertices = 10
Radius = 2, Diameter = 3
Characteristic path length = 1.8
Clustering coefficient = 0

Figure 1. Clustering coefficient

Graphs where vertices represent people and edges represent relationships between people have been studied for more than thirty years; "small-world" networks were originally defined by Milgram (1967). Though published almost a decade later, the earlier work of Pool and Kochen (1978) influenced Milgram as well as our research. Pool and Kochen studied acquaintanceship volume by asking participants in their study to keep a record of individuals with whom they interacted for a period of 100 days. Participants recorded an individual whom they already knew by sight and name when they spoke to that individual. Our definition of when to record an acquaintance relationship was influenced by the definitions given by Pool and Kochen.

The initial evaluation of the social network in the senior class focused on the characteristics that mark a graph as a social network. Many concepts from graph theory are useful in the study of social network graphs, including diameter, radius, center, and eccentricity (Wilson, 1990). Two other characteristics defined in the social network literature, characteristic path length and clustering coefficient (Watts, 1998), are particularly useful to our analysis and are reviewed briefly here.

The *characteristic path length* is a measure of average distance between vertices. It is defined to be the median over all vertices of the mean distance from that vertex to all other vertices. See Figure 1.

The *clustering coefficient of a vertex* is the ratio of the number of edges in the neighborhood of that vertex to the total number of possible edges in the neighborhood. The *clustering coefficient of a graph* is the mean clustering coefficient over all vertices in the graph. Notice that the clustering coefficient is always between 0 and 1. Figure 1 shows two graphs in which the number of vertices, the radius, the diameter, and the characteristic path length are all identical or very similar. The social networks described by these two graphs are quite different, however, and this is captured by the clustering coefficient. A high clustering coefficient from a social standpoint implies tight knit groups with little overlap, while a low clustering coefficient implies lots of overlap but few tight knit groups. An interesting question, sociologically, is to determine what clustering coefficient would be ideal for a social network.

METHODOLOGY

St. Lawrence University, located in Canton, New York, is the oldest continuously operating coeducational institution of higher learning in New York State. It has approximately 2100 undergraduates and a small graduate program in Education (St. Lawrence University, 2003). It is a private, non-denominational, residential, liberal arts college with a strong commitment to both the academic and extracurricular lives of its students.

St. Lawrence University places great importance on the quality of student life and actively seeks input from the students themselves through interaction with student organizations and through student surveys, including an exit survey completed by all graduating seniors. In this atmosphere of information gathering, it made sense to consider capturing information on the social contacts within the senior class.

With a student research grant we set out to capture the acquaintanceship graph for the graduating class of 2002. We leveraged the wired nature of our campus using a Web-based survey that presented students with names and photographs of all 385 students in the senior class.

Our work required defining acquaintanceship between students in the senior class and designing tools to capture that information as painlessly (both for participant and researcher) as possible. Our first thought, surveying the social network literature, was to use multiple (3-5) levels of "friendship" and expect participants to choose among them. Concerns about consistency of definition and the required interface to collect the information led us to abandon this approach.

We settled on a simple, binary definition: an acquaintance is any member of the senior class that you recognize and have spoken to during the current semester. This definition was simple to explain, simple for participants to apply, and led to straightforward data collection software. It was just complex enough to provide an interesting social network for analysis. It closely matched the definition used by Pool and Kochen (1978) where the participant recorded only people that he or she knew and spoke to during the 100 day recording period; the survey was filled out by students approximately 100 days into the Spring semester, matching Pool and Kochen's time frame. Our definition is also similar to that used by Stevenson, et.al. (1997) where a folder could only be passed to another person if the participant knew and had spoken to the individual outside of class at least twice.

People are notoriously bad at estimating the size of their acquaintanceship volume (Poole, et.al., 1978). Our survey addressed this as well as simple forgetfulness through the use of color photographs accompanying senior's names on the survey Web pages. Careful precautions were taken to protect the privacy of our students (all approved by the local IRB): the survey was only available on the campus intranet, every student was given a chance to opt-out of the photographic portion of the survey in two separate e-mailings before the survey went live and through an e-mail link in the live survey. Only six students (1.5%) opted-out; a total of nine students (2.5%) appeared in the survey without photographs for this and other reasons.

The survey was deployed as a series of intranet-only Web pages; Figure 2 shows a (slightly modified) copy of the beginning of the survey web-page. Pages of 50 names/ photos in alphabetic order were generated automatically by our software; navigation was permitted both forward and backward though the finished survey could only be submitted from the last page. Students marked an acquaintance by clicking on the photograph or the checkbox just below it.

When it went live, the survey URL was broadly advertised: campus newspaper, daily lunch bulletins, and the senior class mail and e-mail lists. The e-mails proved particularly effective since the student was reading them on the computer and they

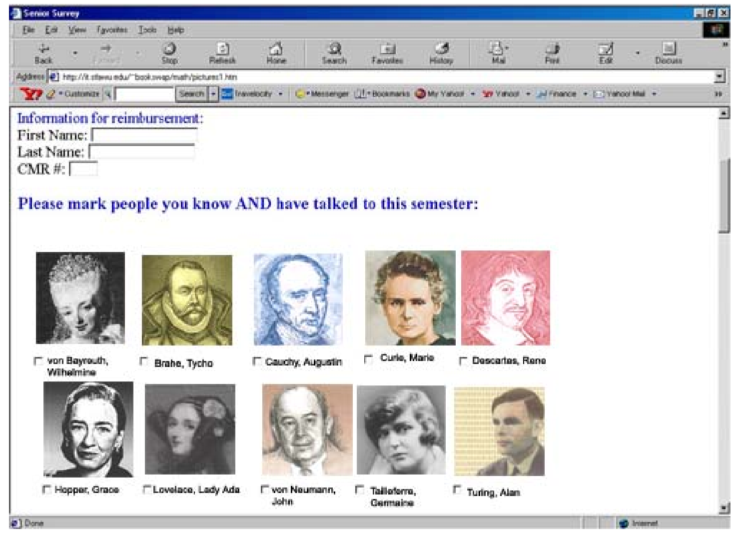


Figure 2. Web page survey

linked directly to the beginning of the survey; students could click and complete the survey in only a few minutes. Each senior who completed the survey received a $5.00 gift certificate from the University bookstore. When the form was submitted, the acquaintanceship data for that student was recorded; participation was limited to seniors with each permitted to fill it out only once. Advertising and ease of use led to a significant participation rate (275 students, 71.4% of the class) and a good representation of the social network of the senior class.

The original data files were transferred offline after the data collection period; they were indexed by student ID numbers and thus contained personally identifiable data. Connecting demographic data compiled by the Institutional Research office with the graph was simple as they were also keyed on student ID number. The annotated graphs were used for institutional analysis and only aggregate information is reported below.

Network Analysis

The first step in processing the raw data was determining whether to model our network as a directed or undirected graph. Given the instructions on the Web survey, an acquaintance relationship *should* be symmetric: any two seniors who had seen and spoken to one another should have checked each other's picture. The majority of reported acquaintanceships were symmetric but a significant fraction were not.

The use of a directed graph (and corresponding definition of two directed clustering coefficients) would have made sure we did not create ties where none existed (if student A recalls an encounter with student B but student B fails to recall it, is student B really acquainted with student A?). Unfortunately it would have also skewed our clustering coefficients; students who failed to answer the survey would have outgoing clustering coefficients of 0. Our use of an undirected graph creates some phantom ties but it permits the inclusion of students who did not fill out the survey. Further work with this data has included some work with directed versions of the graph but there are no significant results to report at this time.

In a class of n = 385 students, the mean size of the acquaintanceship circle was k =113 with a range of 4 to 286. Over the 100 day period, the average senior interacted with 29% of other seniors. The radius of the graph was 2, the diameter only 3. The center of the graph included 193 students, more than half of the senior class. This provides a picture of a highly connected campus where most of the seniors are friends or friends of friends of everyone in the senior class, and where the longest path from any student to any other student is only three links.

The characteristic path length, L, was L = 1.71. The clustering coefficient C for the graph was C = 0.49. On average, half of a senior's acquaintances are, themselves, acquainted. There was a large variation in the clustering coefficient for each vertex, ranging from 0.2 to 0.8, and there was a significant difference between demographic groups.

Demographic Analysis

The University administration found the raw graph statistics for the senior class of 2002 interesting but they had other questions: How does membership in certain groups (*e.g.* international students, minority students, fraternity/sorority houses) impact these statistics; Does our First Year Program (described below) have a lasting impact on a student's social network? Demographic information was provided by the University's Office of Institutional Research and was culled from the registrar (major/minor, GPA, home address) and student entry/exit surveys.

Table 1: Student demographic subgroups

Gender (M / F)
Students of color
Varsity athletes (at least two years)
Member of a fraternity / Member of a sorority
Division of academic major/minor (Science & Math, Social Science, Humanities, Arts, Languages)
Students whose academic major/minor falls into more than one division
Transfer students
International students
High GPA students (over 3.5)
Students from low income families
Students from the area: Northern New York
Students who had studied abroad for a semester or year
Students who were first in their family to go to college
Students from the same first year (freshmen) residence
Students in the same senior residence

Table 1 shows the demographic subgroups we chose to examine. The definition of most of the groups is obvious (a declared major, being on a varsity sports team's roster). The group "students of color" uses the University's standard definition (compiled primarily from admissions data).

In analyzing each group, we looked at both the demographic characteristics of the acquaintanceship circles of the members of the group and at the relative presence of each group in the acquaintanceship circles of other groups. We discuss below the most interesting findings in this analysis.

Gender: The class of 2002 had a gender balance of 54% female and 46% male. We found that females in the class are more well-connected than males, with k_{female} = 116 and k_{male} = 107, although this difference is not statistically significant (p = 0.117). Both males and females interacted more with females, as shown in Table 2. This finding differs significantly from that of Pool and Kochen (1978) in which every group except housewives interacted more with males.

Table 2: Gender in acquaintanceship circles

	Male	Female
Female acquaintanceship circle	40%	60%
Male acquaintanceship circle	47%	53%
Senior class	46%	54%

Females were particularly well represented in the most well connected students on campus, while the least connected students were evenly divided between the genders. The demographic breakdown of the acquaintanceship circles of males and females, based on the groups shown in Table 1, showed few significant differences. Females had a slightly lower clustering coefficient, with C_{female} = 0.48 and C_{male} = 0.51, indicating that on average females interact in a wider variety of circles. The difference in clustering coefficients is significant (p = 0.003).

Students of color: Students of color represented only a small fraction (6%) of the students in the senior class, and we were particularly interested in determining the level of interaction between students of color and the rest of the senior class. Students of color were more well connected than the average student, with k = 117, although the difference was not statistically significant. The average clustering coefficient of students of color was C = 0.45, which was significantly lower than that for other students (p = 0.002), indicating that students of color at St. Lawrence interact in a wide variety of different circles. The acquaintanceship circle of a student of color was remarkably similar demographically to the acquaintanceship circle of an average student, other than including a higher percentage (12%) of students of color. The average acquaintanceship circle for all seniors included 6% students of color, exactly matching the expected value in the population. Furthermore, the average senior knew 29% of all senior students of color, again exactly matching the overall statistic that an average senior knew 29% of all seniors on campus. In short, we found no quantitative evidence in the social network analysis to imply that students of color were isolated from other groups on campus.

Athletes: There is sometimes a perception that student athletes are the center of the campus since they appear in the student paper regularly, win or lose. We were happy to dispel that perception. In our data there was no difference in statistics between athletes and the student body as a whole except for a slight increase in the number of athletes in their social circle (36% v. 31%).

Academic Major/Minor: We divided the academic majors into five divisions: science and math, social sciences, humanities, arts, and languages. We found that the most well connected students, on average, were those majoring in science and math, with $k_{science} = 122$. This average degree for science majors was significantly higher ($p = 0.004$) than the average degree for non-science majors, and was the only academic area for which the difference was significant. Furthermore, the science and math students were well represented in every social group we considered. For *every* demographic group listed in Table 1, the percent of the acquaintanceship circle consisting of science and math majors was higher (and often significantly higher) than the expected value of 38%.

Table 3 shows some of the more interesting findings related to academic major. Not surprisingly, we found that for all divisions the percent of the acquaintanceship circle within that division was higher than the percent expected just from the demographic breakdown. Students naturally interacted more with other students majoring in the same division. However, this increase is most striking for science and math majors. At St. Lawrence University, we promote a great deal of group work, student interaction, and cooperative learning in our science and math classes and labs, and we suspect that these pedagogical techniques may partially explain the higher level of connection between science and math students. Notice from the values of k for different divisions given in the table that these connections seem to be in addition to, rather than in place of, connections with students in other divisions. Our network analysis certainly did not support the stereotype of the isolated science geek.

Table 3: Academic major/minor

Major/minor	% of senior class	average degree, k	% of circle in same division
Science/Math	38%	122	47%
Social Sciences	52%	113	55%
Humanities	36%	111	39%
Arts	10%	114	16%
Languages	5%	115	10%

First Year Program: St. Lawrence University has a long-standing First Year Program that groups first-year students into residential colleges based on a common academic course. Anecdotal evidence has suggested that the friendships formed in these living/learning environments are very strong. Our survey data supports those observations. The average senior, three years later, remained connected with 63% of the members of his or her First Year Program. This finding emphasizes the fact that bonds formed in the first year survive over the four years in college and still play a significant role as a part of a senior's acquaintanceship circle. The connections formed in the First Year Program, however, account for only 17.5% of the total senior acquaintanceship circle, indicating that while students form and keep strong connections in the first year, they also continue to form new connections throughout their time at college.

Transfer students: Our most discouraging findings came with transfer students. The transfer population at St. Lawrence University is small, comprising only 6% of the senior class. The average size of the acquaintanceship circle for this group was significantly lower ($p = 0.000$) than that for all other groups, with $k_{transfer} = 56$. This group also had the highest average clustering coefficient, $C = 0.53$, of any other group except fraternity members. Transfer students were more likely to be female, although other demographic characteristics of transfer students as well as the demographic breakdown of the acquaintanceship circles for these students did not differ in any substantive way from the norm, other than being significantly smaller in size.

Fraternities/Sororities: Members of Greek organizations at St. Lawrence University constituted 38% of the senior class, with 45.6% of all females belonging to a sorority and 29.6% of all males belonging to a fraternity. Members of sororities were remarkably well connected, with $k_{sororities} = 138$, while members of fraternities were closer to the average, with $k_{fraternities} = 119$. Analysis of variance showed significant differences ($p = 0.000$) between all three groups: sorority members, fraternity members, and non-Greek students. A significantly higher than expected number of sorority members were among the most well connected students on campus. There is some evidence that membership in a sorority increases connections with other students while membership in a fraternity can often tend to isolate students. Members of sororities were over-represented in the acquaintanceship circles of almost every demographic group, while members of fraternities were under-represented in almost every group except its own. The average clustering coefficients were significantly different ($p = 0.025$) between the groups, with $C = 0.48$ for sororities and $C = 0.53$ for fraternities.

Characteristics Influencing Connectivity

We conclude our analysis by examining the characteristics that tended to make a senior particularly well connected or particularly poorly connected.

Most connected students: We defined the most well connected students to be the students whose average degree, k, is at least 1.5 times the average, or $k \geq 170$. A total of 54 students fell in this group, or 14% of the class. An average student in this group interacted with 199 seniors over the one hundred day period, or a remarkable 52% of the senior class.

Particularly well represented in this group were female students (61%), members of sororities (44%), students with an academic major/minor in more than one division, science majors/minors, humanities majors/minors, students who studied abroad, and students with a high GPA. Students in this group remained in touch with an amazing 93% of the students in their First Year Program.

The well connected students had low clustering coefficients, ranging from 0.32 to 0.49 with an average of C = 0.42. These students tended to circulate in many different circles.

Least connected students: We defined the least well connected group of students to be the students whose average degree, k, is at most 0.5 times the average, or $k \leq 56$. There are 50 students in this group, or 13% of the class. An average student in this group interacted with only 37 students, or 10% of the class.

Over-represented in this group were non-Greek students, students from low family incomes as well as from Northern New York, first-generation students, and, in particular, transfer students. A full 46% of all transfer students fell into this category. A (non-transfer) senior in this group has maintained contact with only 25% of students in his or her First Year Program.

These students had high clustering coefficients, ranging from 0.49 to 0.85 with an average of C = 0.57.

CONCLUSION

It appears that the St. Lawrence senior class has a "standard" social network, a graph with small L for all random graphs of its size and a high C. This network is represented by a "small world" graph (Watts, 1999). We have learned that seniors at St. Lawrence University tend to be very well connected and we were pleased to observe how diversified their acquaintanceship circles actually are. The analysis presented in this paper, however, is only the beginning. We hope to apply group/clique analysis to this data to validate both the method and our methodology. With our extensive demographic information on students, it will be interesting to see how the group analysis provides additional insight into, for example, various Greek houses or athletic teams.

This paper reports the first of what we hope are many senior class social network surveys at St. Lawrence University. As we gather more data we will be able to measure the differences in social network statistics between the various senior classes, and how connections change over time. We hope to be able to measure quantitatively the effectiveness of efforts made by the University to better integrate groups of students, such as transfer students. We also hope to coordinate the information in this data set with the data we obtain from senior class exit surveys that measure various dimensions of student satisfaction with the academic and social life at St. Lawrence.

Our research addresses the lament cited at the start of this paper that too many methods chase too few social network data by providing another large data set. We plan to continue addressing it by surveying future graduating classes. The analysis presented here is a jumping off point for more sophisticated examinations of the data. The significance of the clustering coefficient is a question raised rather than answered by this work. Our work is mathematical, examining characteristics of the social network data. Several sociological questions arise: What sort of statistics would the graph of the social network at an "ideal" university have? What do different clustering coefficients tell us about the social climate? We have just begun to wrestle with these questions.

REFERENCES

Guare, J. 1990. *Six Degrees of Separation: A Play.* New York : Vintage Press.

Martin J, Yeung K, and Zablocki B. 2001. The Urban Communes Data Set: A Gold Mine for Secondary Analysis, *Connections* 24: 54 – 59.

Milgram S. 1967. The Small World Problem, *Psychology Today* 1: 60 – 67.

Emily Eakin. 2003. Connect, They Say, Only Connect. *New York Times*, January 25, 2003.

Pool I and Kochen M. 1978. Contacts and Influence, *Social Networks* I: 5 – 51.

Reynolds P. 1999. *The Oracle of Bacon*, http://www.cs.virginia.edu/oracle/.

St. Lawrence University, 2003. *About SLU*, http://www. stlawu.edu/glance.html/.

Stevenson W, Davidson B, Manev I, and Walsh K. 1997. The Small World of the University: A Classroom Exercise in the Study of Networks, *Connections* 20: 23 – 33.

Watts D and Strogatz S. 1998. Collective Dynamics of 'Small-World' Networks, *Nature* 393: 440 – 442.

Watts D. 1999. Small Worlds: The Dynamics of Networks Between Order and Randomness. Princeton: Princeton University Press.

Wilson R and Watkins J. 1990. *Graphs: An Introductory Approach.* New York: John Wiley and Sons.

CONNECTIONS 27(2): 59-76

http://www.insna.org/Connections-Web/Volume27-2/Takacs.pdf

Effects of Network Segregation in Intergroup Conflict: An Experimental Analysis[1]

Károly Takács[2]
University of Groningen

Dense in-group and scarce out-group relations (network segregation) often support the emergence of conflicts between groups. A key underlying mechanism is social control that helps to overcome the collective action problem within groups, but contributes to harmful conflicts among them in segregated settings. In this study, a new experimental design is introduced to test whether internalized social control affects contribution decisions in intergroup related collective action. Subjects played single-shot Intergroup Public Good games in two groups of five without communication. Subjects were connected via computers and connection patterns were manipulated to detect forms of social control that are activated conditional on expectations and on the composition of the artificially created ego-network. Results confirm the influence of behavioral confirmation and the conditional impact of internalized selective incentives. As an aggregated consequence of these social control effects, harmful intergroup outcomes were least likely when members of the groups were arranged in a mixed network.

INTRODUCTION

Single-shot social dilemma experiments consistently find nonzero cooperation rates. A lot of people act against their egoistic interests and make sacrifices for the collectivity also in strictly impersonal settings in which no communication is allowed and subjects are completely strangers to each other. In a competition situation with another group, experiments find even higher contribution rates to the provision of a public good (Bornstein, Erev, and Rosen, 1990; Schopler and Insko, 1992; Bornstein and Ben-Yossef, 1994; Insko et al., 1994; Bornstein, Winter, and Goren, 1996). When intense intergroup competition leads to negative consequences for members of both groups, public "bads" are provided instead of public goods. Why do people still act in favor of their groups under such circumstances?

This paper argues that the monetary payoff structure of experimental games does not fully describe the incentives of subjects in the laboratory. The emphasis here will be on the role of incentives that stem from interpersonal relations and social networks. The importance of social networks in social dilemmas was highlighted by both theoretical (e.g., Marwell, Oliver, and Prahl, 1988; Gould, 1993; Flache and Macy, 1996; Chwe, 1999) and empirical studies (e.g., McAdam, 1986; Chong, 1991; Finkel and Opp, 1991; Gould, 1995; Sandell and Stern, 1998). Network effects are attributed to the fact that individuals are influenced by the presence, opinion, expectations, and behavior of friends, neighbors, colleagues, and relevant others, when they decide to participate in collective action. These mechanisms can be summarized as social control (cf. Kornhauser, 1978; Gibbs, 1981; Black, 1984; Heckathorn, 1990; 1993; Macy, 1993; Villareal, 2002).

Only a limited amount of studies have tried, however, to describe and measure these effects in a controlled environment (some indications are given for the presence of social control by Yamagishi, 1986; van de Kragt, Dawes, and Orbell, 1988; Rapoport, Bornstein, and Erev, 1989; McCusker and Carnevale, 1995; Gächter and Fehr, 1999; Rege and Telle, 2004). Structural considerations were disregarded by previous experiments on intergroup relations (an exception is Grö βer and Schram, 2006). In general, the experimental literature that takes account of networks is limited but growing (for an overview, see Kosfeld, 2003). These avenues

[1] Address of correspondence: Interuniversity Center for Social Science Theory and Methodology / Department of Sociology, University of Groningen, Grote Rozenstraat 31, 9712 TG Groningen, the Netherlands. E-mail: k.takacs@rug.nl. The paper was revised while the author was a junior fellow at the Collegium Budapest / Institute for Advanced Study and a Magyar fellow at the Netherlands Institute for Advanced Study, Wassenaar.

[2] I am grateful to Roel Popping, Frans Stokman, Andreas Flache, and Wim Liebrand for their valuable and thorough advises, comments, and criticisms. I thank Anthony Oberschall, John Dovidio, Marcel Lubbers, Carlijne Philips, Rafael Wittek, Marcel van Assen, Brent Simpson, Alexander Gattig, Christian Steglich, René Thorenvlied, Corine Hoeben, Uwe Matzat, Béla Janky, Marieke van der Wal, Rita Smaniotto, Martin van der Gaag, Alinda van Bruggen, Jan Kratzer, and reviewers for their comments on the experimental design and/or on an earlier version of this paper. I also would like to thank Sicco Strampel, Jeanette Loos, and Rie Bosman for their practical help in conducting the experiments.

should be pursued to gain further insights into determinants of individual behavior in social dilemmas. This paper argues that social control in certain forms and also elementary structures might be present in the laboratory and can make a significant difference to contribution decisions, even when subjects do not know each other and are not allowed to communicate.

As a model of intergroup relations, an extension of the Intergroup Public Goods (IPG) game (Rapoport and Bornstein, 1987; Takács, 2001) will be used that represents the dichotomy of interdependencies within the groups (provision of a public good) and between the groups (intergroup competition for a scarce resource). In this game, players are divided in two groups. Every player can decide to contribute or not to the provision of a public good. Contribution is costly. The number of contributors is compared between the groups. All members of the group with more contributors receive a public good reward v and all members of the other group receive a "public bad" d. In case the number of contributors are equal, all players receive a punishment reward c (v>0>c>d). The IPG game in this form is intended to model group competitions such as civil war, conflicts between pupil groups, fights between football supporters or urban gangs.

In case of only few contributors, nothing happens, the status quo is preserved. It means that if both groups have less contributors than a minimal contributing set (MCS), no public good or bad is provided (cf. van de Kragt, Orbell, and Dawes, 1983). In this paper, an outcome will be called intergroup conflict, if one or both of the groups receive negative public rewards (c or d), or equivalently, the number of contributors at least in one group is above the threshold (MCS). Assuming no other incentives, the outcome of this game should not depend on the network connections group members might have between each other.

In order to capture relevant network effects, the IPG model has been extended by assuming dyadic mechanisms of social selective incentives and behavioral confirmation (Takács, 2001). These forms of social control have been shown to be possible underlying mechanisms why social networks might influence the likelihood of intergroup conflict. The extended model predicts that in particular, network segregation affects the likelihood of intergroup conflict and the relationship can be characterized by an S-shape function. This implies that segregation is likely to promote intergroup conflict, but in extreme ranges of segregation, an additional change does not result in an increase in the likelihood of conflict (Takács, 2001). These theoretical predictions directly lead to the main question and hypothesis of this study. In the context of a laboratory environment, is intergroup conflict indeed more likely when group members are arranged in a segregated network?

SOCIAL CONTROL AND NETWORK EFFECTS IN EXPERIMENTS

This study will examine what forms of social control back the effect of network segregation on intergroup conflict, if there is any. It will be explored in controlled experimental conditions what forms of internalized social control influence the decision of subjects to contribute or not to the provision of intergroup public goods.

The following fundamental forms of social control will be considered as possible mechanisms. In-group social selective incentives, such as prestige, respect, and status either reward those who contributed to the group welfare (e.g., Lovaglia, Willer, and Troyer, 2003) or punish those who did not make contributions. Empirical studies show that social selective incentives are disseminated mainly locally, through interpersonal relations (Sandell and Stern, 1998) and are often internalized as contribution norms that create a cognitive reward for cooperation (Scott, 1971; Kornhauser, 1978; Coleman, 1990: 293). Individuals feel rewarded when they "did the right thing for the group" (Opp, 1989).

A similar form of social control is present in network relations with out-group members. Members of the competing groups, however, have contradictory interests in intergroup competition and therefore they reward each other's action that is against the in-group interest (e.g., Kuran, 1995, 9-10). These relations therefore transmit social selective incentives that punish contribution and reward defection. Out-group selective incentives are also likely to be internalized as a fear from dyadic conflict and benefit for local harmony. Their relevance can provide an explanation why contact can help to normalize intergroup relations (cf. Allport, 1954).

Another prominent form of social control is behavioral confirmation that expresses the desire to conform to the expected behavior of relevant individuals. It means that doing the same as relevant others has a positive value by itself and increases the utility of both sides independently from future interactions. In empirical collective action situations (e.g., strikes, demonstrations, and revolutions) participation in collective political action can be largely explained by willingness to conform to the behavioral expectations of relevant others (e.g., Finkel and Opp, 1991; Chong, 1991; Oberschall, 1994). There is indication for the relevance of such a mechanism also in public good experiments (Yamagishi, 1986; McCusker and Carnevale, 1995; Rege and Telle, 2004). Behavioral confirmation has a two-fold effect: confirmation by participating in-group alters provides an incentive for contribution and confirmation by free riders works against contribution. Even if others are not able to monitor individual choice, behavioral confirmation might affect decisions as an internalized mechanism or imitation strategy (Asch, 1956; Dawkins, 1976; Pingle, 1995).

As an aggregated consequence of dyadic social control, the network structure of individual relations influences the likelihood of intergroup conflict. Dense in-group relations and scarce out-group relations are correlated with extensive distribution of social selective incentives between in-group members and limited realization of out-group selective incentives. Hence, network segregation supports contributions to harmful intergroup competitions and consequently to the emergence of harmful conflicts. The underlying mechanisms responsible for this are the fundamental forms of social control.

A major difference compared to field situations is that subjects are unknown to each other in the laboratory; consequently there are no social network relations between them. Can social control operate under such circumstances?

Experimental evidence shows that face-to-face contact facilitates cooperation in conflict situations (cf. Drolet and Morris, 2000). Previously, this finding was explained by the social psychological process of rapport that is conceptualized as a "state of mutual positivity and interest that arises through the convergence of nonverbal expressive behavior in an interaction" (Drolet and Morris, 2000: 27; Tickle-Degnen and Rosenthal, 1990). There is no doubt that when subjects are able to communicate with nonverbal signs or are able to send emotional signals, they influence the behavior of each other in the social dilemma task. The question is whether minimal contact and a "minimum network" have an additional effect that is due to the activation of internalized social control.

HYPOTHESES AND EXPERIMENTAL DESIGN

Minimal contact and social control

To test the presence of different forms of social control and the segregation effect on intergroup conflict in a controlled environment, a new experimental design is introduced. In the experiments, the seating arrangement of subjects and visibility conditions were manipulated in order to detect forms of social control that are activated conditional on the composition of the ego-network that is created experimentally. Minimal contact was introduced between connected subjects in the form that subjects were able to see to whom they are connected and they were able to identify the group membership of each other. Verbal and nonverbal communication was disallowed to avoid application of other forms of social control and signaling. It was tested whether this minimal contact is sufficient to activate internalized forms of social control.

In later parts of the experiments, additional to minimal contact, monetary side-payments were introduced as representations of external behavioral confirmation and in-group selective incentives. These effects are expected to be stronger than internalized effects. With their introduction a meaningful comparison can be made between the size of monetary and internalized social control. With regard to forms of social control, the following hypotheses are explicated.

Selective incentives: In-group selective incentives have a positive effect on contribution propensities. More connections to members of the in-group mean the distribution of selective incentives from multiple sources. Hence, the higher the number of in-group members in the ego-network, the higher the contribution rate is.

The presence of contacts to members of the opposite group triggers a similar, but opposite effect. Internalized out-group selective incentives have a negative effect on contribution propensities. The higher the number of members of the opposite group in the ego-network, the lower the contribution rate is. Because of its similarity with in-group selective incentives, this form of social control was not introduced in a monetary form in the experiments.

The effect of behavioral confirmation is not only dependent on the composition of the ego-network, but also on expected decisions of alters. It is presumed that subjects do not make qualitative differences between alters who are members of the same group.

Behavioral confirmation of in-group members: Behavioral confirmation is predicted to have an effect on contribution propensities. The direction and the size of the effect depend on the number of expected contributors and on the number of expected defectors in the ego-network. If the former is higher, the effect is positive. If the latter is higher, the effect is negative. It is assumed that the size of the effect is a linear function of the difference between the two.

For the operationalization of behavioral confirmation, the expectations of subjects were measured by asking them to forecast the decision of their left and right neighbors before every decision round.

Network segregation and experimental implementation

Network connections are conceptualized as adjacency in the seating configuration in the experiment. As neighbors are expected to be the direct source of social control, different neighborhood compositions would lead to different contribution propensities. At the aggregated level, different outcomes can be predicted for different neighborhood structures. From the nature of the specified social control mechanisms it follows that segregation is likely to promote intergroup conflict (cf. Takács, 2001). On the basis of this theoretical prediction, the following hypothesis can be formulated for the IPG experiments:

SEGREGATION HYPOTHESIS: *In a segregated structure, contribution rates will be higher and intergroup conflict will be more likely than in a mixed structure or in a control condition with no networks.*

Furthermore, Takács (2001) also specified the impact of the relative size of social control mechanisms on intergroup conflict. As in-group selective incentives always drive towards contribution and behavioral confirmation might drive towards contribution as well as towards defection, the segregation effect on intergroup conflict is stronger where in-group selective incentives are relatively important when compared to behavioral confirmation. In order to test this theoretical prediction, in one experimental condition external in-group selective incentives and in another experimental condition external behavioral confirmation were introduced as additional monetary side-payments. On the basis of the theoretical prediction, the hypothesis about these effects is as follows:

The segregation effect on the likelihood of intergroup conflict will be stronger in the monetary selective incentives condition than in the monetary behavioral confirmation condition.

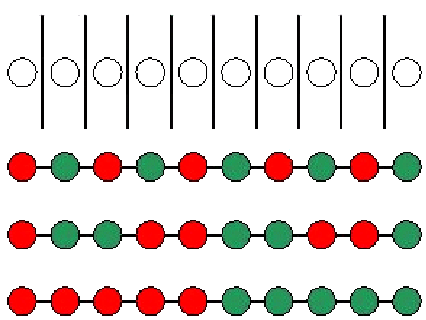


Figure 1 Structural Conditions in the Experiments: Control Condition, Low, Medium, and High Segregation. Note: red and green nodes indicate members of red and green group. In the Control Group no color labels were introduced and panel walls separated the subjects.

To test the above hypotheses, three types of network arrangements were implemented between sessions; with low, medium, and high segregation (see Figure 1). In addition, every experiment started with a control condition, in which subjects made their decisions in isolation without the knowledge of their group membership. After the control condition, color labels indicating group membership were introduced and subjects were arranged in one of the network conditions (low, medium, or high segregation) that are shown in Figure 1. For instance, in the case of low segregation, all subjects in the red group were seated next to members of the green group.

Subjects could see the composition of their ego-network on their computer screen. This intervention is targeted to assess internalized social control effects in the presence of minimal contact.

The IPG game and experimental implementation

The experiment used a series of single-shot IPG games as a model of competitive intergroup relations. The payoffs of the game used in the experiment are outlined here. There were two groups: the red group and the green group consisting of five members each. Every player had to decide individually whether to keep a bonus of 11 points completely (1 point was equivalent to approximately 0.42 USD) or to give all of it to help their group in the competition. Depending on the number of contributors in the groups, public good and "bad" rewards were distributed equally among all group members. The sizes of these rewards in the experiments are shown in Figure 2. Each member of the group with more contributors received 15 points and each member of the group with less contributors lost 15 points as long as there were at least three contributors in the winning group (v=15; d=-15; MCS=3). A minimal contributing set with three persons was chosen in order to avoid that few coincidental contributions would have affected the result and in order to decrease individual efficacy in the experiment. Less than three contributions were insufficient to produce a public good and these contributions were lost to these individuals. When the number of contributors was equal in the groups and was over the minimal contributing set, all subjects lost 11 points (c=-11).

Figure 2. *The IPG Game Used in the Experiments*

payoffs in points		*number of contributors in the green group*					
		0	**1**	**2**	**3**	**4**	**5**
number of contributors in the green group	**0**	0	0	0	15	15	15
		0	0	0	-15	-15	-15
	1	0	0	0	15	15	15
		0	0	0	-15	-15	-15
	2	0	0	0	15	15	15
		0	0	0	-15	-15	-15
	3	-15	-15	-15	-11	15	15
		15	15	15	-11	-15	-15
	4	-15	-15	-15	-15	-11	15
		15	15	15	15	-11	-15
	5	-15	-15	-15	-15	-15	-11
		15	15	15	15	15	-11

Note: The payoffs are public good rewards distributed to everyone in the red (bottom left corner of each cell) and in the green (top right corner) group. In addition to these payoffs, defectors could keep the endowment of 11 points, and every subjects received 15 points to ensure positive payments.

Everyone received these public good and "bad" rewards, regardless of the decision to keep or give away the bonus of 11 points. Figure 2 does not include the bonus reward that is added to the payoff of those subjects who decided to keep the bonus. Moreover, to ensure positive payoffs, every subject was entitled to an additional payment of 15 points at the end of the experiment.

In order to obtain more reliable data in the experiments, the game was played many times in each session, but subjects received payments in a randomly selected single round only. No information was provided during the experiment about what has happened in earlier rounds and what others were doing in the same round. In this way, every decision round could be handled in an equivalent way. This method was applied in earlier team game experiments by Bornstein and Ben-Yossef (1994).

Table 1. Overview of Experimental Parts

Part I	*anonymous control condition*	*panel walls isolate subjects*
Part II	*minimal contact is established*	*networks (low / medium / high segregation)*
Part III	*one form (b/s) of social control is introduced in a monetary form*	*networks (low / medium / high segregation)*
Part IV	*the other form (s/b) of social control is introduced in a monetary form*	*networks (low / medium / high segregation)*

Every experiment consisted of four parts (see Table 1). In Part I, subjects made their decisions in isolation. In Part II, subjects played the IPG game with minimal contact in different network configurations that are represented in Figure 1.

Table 2. The Number of Sessions by Experimental Conditions

level of segregation	low	medium	high
monetary behavioral confirmation is introduced in Part III	*3*	*4*	*3*
monetary in-group selective incentives introduced in Part III	*3*	*4*	*3*

The comparison of contribution rates in Parts I and II will provide the opportunity to test the main hypotheses about the presence of internalized social control mechanisms and segregation effects. In Part III, monetary side-payments were introduced between connected subjects. This intervention aimed to provide a meaningful comparison of the relative size of the effect of internalized social incentives and monetary side-payments. With regard to monetary side-payments, two conditions were implemented between experimental sessions. Next to the payoffs that were present in the beginning of the experiments (see Table 2), in Part III, in the monetary behavioral confirmation condition external behavioral confirmation incentives (5 points), in the monetary selective incentives condition external in-group selective incentives (5 points) were introduced (cf. Table 1). In Part IV, in both conditions the other external incentives were also introduced. Subjects received 5 points of behavioral confirmation reward if one of their in-group neighbors chose the same action as they did and received 10 points if two of their in-group neighbors acted the same way. In-group selective incentives were distributed regardless of the decision of neighbors. Contributors received 5 points for each in-group neighbor they had. Out-group selective incentives were not introduced in a monetary form. In the low segregation condition (six sessions) there was no change due to the absence of in-group neighbors and this condition was used as a control condition. To summarize, the experiment has followed a 2×3 block-design that is represented in Table 2.

The order of experimental parts shown in Table 1 was not altered, since once identities are assigned to subjects there is no logical way back to a no-identity treatment. The design is therefore not perfectly counterbalanced, and results have to be interpreted with the reservation that control for ordering effects was not possible.

Experiments were combined with repeated IPG games. Repeated games followed single-shot games in all four experimental parts. Experiments were designed so as to exclude possible influences of previous decisions. Subjects were explicitly told before every part that previous parts and repeated games are completely independent from the next part. New parts always started after a short break and with introductory instructions that attempted to create the impression as if nothing has happened before in the experiment. This manipulation, however, cannot perfectly exclude the possibility of history effects that will be discussed later among control variables.

METHOD

Subjects

203 subjects took part in the experiments at the University of Groningen, in the Netherlands. Subjects were recruited via e-mail and board advertisements promising monetary rewards for participation. All 203 subjects completed the decision tasks and only two have failed to complete the post-decision questionnaire. Altogether, 21 sessions took place and subjects made 4060 single-shot game decisions (20 each). The intended number of participants was ten in all the 21 experimental sessions. On average, thirteen subjects were invited to the sessions as it was anticipated that some would not show up. Four sessions failed to be completely filled. In these cases, computer players were included.[3] Subjects were told that they are programmed in a way to resemble human behavior. In fact, they were simple programs playing mixed strategies with condition-dependent probabilities of contribution. Human decisions in the incomplete experiments are also included in the analysis, but computer decisions are excluded. The inclusion of simulated partici-

[3] This meant 1, 2, 2, and 2 cases in these four sessions.

pants did not have a significant influence on the behavior of subjects in the IPG games.[4]

114 (56.2%) subjects were female. 187 (92.1%) subjects were university students at the time of the experiments and 16 had already graduated. Students came from all faculties of the university: 55 studied behavioral or social sciences, 47 subscribed for literary studies or art, 26 studied natural sciences, 17 studied law, 13 studied economics, 10 were students at the business faculty, there were 8 students of medical science, 8 subjects studied spatial sciences, and one subject read philosophy. Because of similarities and for the sake of simplicity, economic, business, and spatial sciences were merged in the analysis (furthermore, these faculties have the same physical location) and the student of philosophy was allocated to the category of literary studies and art. The college major of two subjects was unknown.

Single-shot games (only the decision rounds) took approximately three minutes in each experimental part. During this time subjects had to make five decisions. The entire experiment was on average 80 minutes long.

The payoff for subjects was contingent on their decisions, as well as on the decisions of other participants of the session. Individual payoffs were calculated on the basis of outcomes in the single-shot and in the repeated games. From the single-shot games, only one was selected randomly in each experimental part to be included in the calculation. This payoff had a weight of five rounds (the number of single-shot games in one experimental part). Total payoffs varied between 14 and 32 points with an average of 21.1 points that was equivalent to 8.9 USD. If subjects ran out of decision time, a random decision was implemented with 50% chance of contribution. For all such cases, final payment was decreased by 1%. This happened only 26 times out of 4060 decisions (0.64%). Random decisions are not included in the analysis.

Procedure

Experiments were conducted in the same computer laboratory. Upon arrival, subjects were randomly seated at a computer.[5] Panel walls separated the subjects to ensure their privacy. Subjects received instructions on paper and on their screen.[6] After reading the instructions they were allowed to ask the experimenter questions. After the questions had been answered, subjects were not allowed to talk. All participants strictly adhered to the rules. After the questions, an examination of understanding followed.

In each of the four experimental parts, subjects played five rounds of single-shot IPG games, and a randomly chosen number of repeated games afterwards. In every decision round, subjects had to decide whether they would keep the 11 points bonus or give it to help their group to achieve success in the competition. These two options appeared in a randomized order on their screen. The bonus was represented also graphically as a bag of money. Subjects were assured of the anonymity of their decisions and that they would receive the amount of money they earned during the experiment in sealed envelopes, after the experiments had ended. In the single-shot games, it was announced that every decision counts towards the final payment, but that only one game of each part would be chosen randomly for payment.

In the beginning of Part II, panel walls were removed and group membership was made public by the experimenter. Red and green flags were attached to the monitors and subjects also received an A-4 colored paper with the color of their group. In each condition, subjects were arranged behind computers due to the neighborhood configuration of the given session. Participants could clearly see the indication signs of group membership of their neighbors, and with some effort they could also check membership of more distant subjects. Subjects played five rounds of the same IPG game again. Before every decision in Part II, III, and IV, subjects had to give their expectations about the subsequent decision of their neighbors. The five single-shot games were followed by repeated games in each part.[7]

Calculation and announcement of the individual results followed the experiment. Meanwhile subjects were asked to fill in a questionnaire on their computer. Monetary payments were supplied in sealed envelopes. The first subject, who had completed the questionnaire, could go immediately to the experimenter to receive payment. Other subjects had to wait until they got a signal from the server. Hence, subjects left the laboratory individually, with a short time difference between their departures.

ANALYSIS OF CONTRIBUTION PROPENSITIES

This section describes the logic of analysis that is used to test the main hypotheses. Besides the main effects of social control that are believed to be the underlying mechanisms of the segregation effect on intergroup conflict, the influence of personal characteristics are discussed that are handled as control variables.

For the analysis of experimental data multilevel logistic regression is used (Bryk and Raudenbush, 1992; Goldstein, 1995). There are two levels in this case. Single decisions are the lower level observations and subjects, who took these

[4] A group-level control variable indicating the presence of a computer player was not significant when added to any of the multivariate models discussed in the Results section.

[5] The computer program for the experiment was written by Sicco Strampel in Delphi.

[6] Full instructions are available in Takács (2002: 101-104).

[7] In the repeated games, subjects were informed about the result of the previous round.

decisions, and their characteristics are the higher level observations. The two-level model corrects for the methodological problem that observations within the subjects are not independent. Multilevel models take care of this dependency and separates within subject and between subject variance. For the binary dependent variable of individual contribution, the logit transformation is used. Formally, let the function Pri denote the propensity of actor i to cooperate in the rth single-shot game. Note that while the probability of contribution is between 0 and 1, the propensity can take any value. The propensity of cooperation is specified by the logit link function (Goldstein 1995: Chapter 7), which is the natural logarithm of the quotient of the probability of contribution Pri(C) and the probability of defection Pri(D):

$$P_{ri}^{I} = \ln\left(\frac{P_{ri}(C)}{P_{ri}(D)}\right) = \alpha_0 + \varepsilon_i + \xi_{ri} \qquad (1)$$

where α_0 is the baseline contribution propensity. Notation ε_i stands for a subject level error term and ξ_{ri} is intra-individual variation. The latter term represents the residual variance that is not estimated in models that include the random intercept α_0. It is assumed that the subject level error has a zero expected value and has a normal distribution, formally

$$\varepsilon_i \sim N(0,\sigma^2)$$

where the variance σ^2 is going to be estimated. This baseline model does not contain any explanatory variables and allows to model behavior in the anonymous control condition (Part I).

Intra-individual variation results from experimental manipulations. These main factors are relevant after the introduction of minimal contact in Part II. Additional reasons for intra-individual variation that can already be present in the control condition are stochastic individual decisions, consideration of mixed strategies, or simply inconsistency. Since nothing distinguishes between single-shot game rounds, only a low intra-individual variation is expected within an experimental part that might be due to individual uncertainty or inconsistency. In the simplest model, it is assumed that intra-individual variation is not correlated with round number r and has a zero expected value. However, this assumption will be relaxed and a trend element will be added, if there are indications of learning the structure of the game through the experiment.

Main effects: social control

With the introduction of minimal contact and network structures (Part II), the effect of segregation on intergroup relations and the presence of underlying internalized mechanisms can be tested. The number of in-group ties is predicted to have a positive effect on contribution rates as minimal contact allows for the activation of internalized in-group selective incentives rewarding contribution and punishing defection (s_0). The number of out-group ties is predicted to provide an incentive against contribution because of internalized out-group selective incentives (t_0). An auxiliary assumption here is that internalized selective incentives affect contribution propensities as a linear function of the number of ties. This number varies between subjects; it is zero in Part I for all subjects and might be 0, 1, or 2 in later parts of the experiment depending on the network condition.

The expected behavior of in-group alters is relevant for internalized behavioral confirmation as subjects are predicted to adjust their actual decisions to the expected decision of alters from their group. The difference between the expected number of contributing in-group alters and the expected number of defecting in-group alters is predicted to have a positive effect on contribution rates (captured by the parameter b_0). As it was expressed earlier, if the expected number of contributors is higher, behavioral confirmation increases the likelihood of contribution. On the other hand, in case there are more defectors among alters, behavioral confirmation decreases the likelihood of contribution. Because of the simple network patterns used in the experiment, the difference can only take integer values between -2 and 2.

The parameter values of s_0, t_0, b_0 are estimated from the experimental results. The relative weight of the utility of monetary rewards and of the utilities attached to different forms of non-monetary incentives can change from person to person. Therefore, no specific form of utility function is assumed that could be applied to everyone. In the simplest model, only the average individual importance of internalized social control is estimated, but some presented models will allow for a random variance in the size of these effects. Models with random effects will assume that the effects of internalized social control for the subjects are normally distributed around their means. This is consistent with the statement that individuals do not assign the same relative utility for social control, but the utilities are scattered normally around a certain mean evaluation. In this part of the analysis, variances of the effects of different forms of internalized social control will be estimated, as well as their covariances.

For a better calibration of social control effects, in some sessions from Part III on, external social control is introduced in the form of monetary side-payments. External selective incentives (s) and behavioral confirmation (b) are predicted to have a positive effect on contribution rates and these parameters are also need to be estimated. Effects of external social control can clearly be separated from internalized social control, as in Part II of the experiments only internalized social control could have an effect. The size of the effect of external control, however, might interact with the size of the effect of internalized social control. In general, the utility of monetary rewards might differ subject by subject, therefore, part of the multilevel analysis will allow for a random variation in their sizes over the subjects.

Control variables and interaction effects

Previous experiments revealed several important factors that influence cooperation rates in social dilemmas (e.g., Ledyard, 1995). The inter-individual variation of contribution propensities in intergroup related collective action might also depend on personal characteristics, like gender, college major, experience in similar experiments, attitudes towards risk, or social orientations. These factors will be included in the analysis as control variables; therefore no hypotheses are explicated about their effects. They are included as controls because they enrich research with interesting insight and comparisons can be made with previous findings.

For instance, there are contradictory findings in previous social dilemma experiments about whether women or men are more cooperative (e.g., Isaac, McCue, and Plott, 1985; Stockard, van de Kragt, and Dodge, 1988; Mason, Phillips, and Redington, 1991; Frank, Gilovich, and Regan, 1993; Brown-Kruse and Hummels, 1993; Nowell and Tinkler, 1994; Cadsby and Maynes, 1998; Eckel and Grossman, 1998; Ortmann and Tichy, 1999). Most subjects participating in experiments are students at different faculties of the university. Direction of study might cause individual differences in willingness of contribution. Previous research found that economists have lower contribution rates (Marwell and Ames, 1981; Carter and Irons, 1991; Frank, Gilovich, and Regan, 1993), although there are also experiments that do not find this effect(Isaac, McCue, and Plott,1985;for an overview, see Ledyard,1995:161, 179).

Besides these background variables, relevant factors include attitude measures that indicate special forms of individual utility functions. Previous findings show that attitudes towards risk correlate with contribution propensities (Suleiman and Or-Chen, 1999). Since the contribution decision involves the possibility of a higher reward, but also involves the risk of losing the bonus completely, subjects with a risk-seeking attitude might have higher contribution rates (Budescu, Rapoport, and Suleiman, 1990). On the other hand, there are arguments that in repeated social dilemmas risk aversion increases cooperation (Raub and Snijders, 1997; van Assen and Snijders, 2002). In the experiments of this study, attitudes towards risk were included only as control variables. For the measurement of risk preferences, questions with preference comparisons (see Farquhar, 1984) were used.

Utility functions can also include altruistic elements, which certainly influence rational decision-making in social dilemma experiments (e.g., Liebrand, 1984; Doi, 1994). Subjects, who order positive utilities for the gains of others, behave differently from individualistic ones. For the approximation of such utilities, standard questions regarding social orientations were used. They consisted of a series of decomposed games with an unknown person.[8] The measurement presumed that individuals are only *prosocial* (cooperative), *individualistic*, or *competitive*. Previous research found only these types relevant in describing human behavior (van Lange et al., 1997; van Lange, 1999; Suleiman and Or-Chen, 1999). Among each type an egalitarian tendency was distinguished (cf. van Lange, 1999). Although in a two-person PD game or in a public good experiment higher contribution rates are expected from prosocial subjects, it is not at all evident in the IPG game. One could argue that subjects who order utility weights for rewards of unknown others, would do this equally for everyone, including out-group members. Consequently, their contribution rates would not be different from individualistic subjects. A counter-argument is that prosocial (and also egalitarian) orientation is associated with high utility for social identity, which is obtainable in a relational comparison with the out-group. Hence prosocial orientation is primarily directed towards in-group members. Results will show whether prosocial individuals are more concerned about harmful outcomes and thus abstain from contribution or whether they have higher contribution propensities and are even the initiators of harmful intergroup conflict.

Some of the participants knew each other. As acquaintances might influence actual decisions in the experiment, the number of acquaintances in the experiment is included as a control variable. In part of the analysis, interaction effects of background variables and social control are also included, because the relative size of internalized social control in the utility function might depend on certain personal characteristics. There are contradictory findings in previous experiments about whether people are more likely to think of others of the same sex to be contributors and in general, whether men or women are more likely to be thought of as better contributors (Ortmann and Tichy, 1999; Solnick and Schweitzer, 1999). For explorative reasons, interactions between gender and social control and interactions between social orientations and social control are also included as control variables.

Since experiments were designed to separate motives in single-shot situations from incentives that are present in repeated play, no history effects are expected on single-shot decisions, but as a test of this hypothesis, previous outcomes of iterated games were included as control variables in part of the analysis.

[8] The exact questions can be found in Takács (2002).

Table 3. Outcomes by Segregation Conditions in the Experiments

segregation condition in the experiment	*outcome of the decision round* *no competitive action*	*conflict*	Total
control condition (unknown group membership)	*26.97% (271)*	*73.03% (734)*	*100% (1005)*
low segregation	*50.23% (428)*	*49.77% (424)*	*100% (852)*
medium segregation	*13.75% (160)*	*86.25% (1004)*	*100% (1164)*
high segregation	*11.85% (120)*	*88.15% (893)*	*100% (1013)*
Total N	*24.27% (979)*	*75.73% (3055)*	*100% (4034)*

Note. Cases in parentheses are weighted (multiplied) by the number of human decisions in the given game. For the χ^2 test unweighted outcomes are used, N = 420

RESULTS

Contribution rates and conflict under different experimental conditions

As the consequence of dyadic social control, different outcomes were expected by segregation conditions. The segregation hypothesis predicted that conflict is least likely in the mixed condition and is most likely in the highly segregated setting. Table 3 summarizes the experimental outcomes by segregation conditions. The hypothesis that the outcomes of the IPG game are independent of segregation conditions can be rejected ($\chi^2(3)=46.370$, $p<0.001$).

Table 3 shows that conflict was already quite likely in the control condition. It indicates that many subjects have contributed even when they were isolated, which cannot be explained by social control effects. Conflict was much less likely in the low segregation condition, and occurred most often in the high segregation condition, which supports the segregation hypothesis. On the other hand, conflict was almost as likely in the medium segregation condition as in high segregation. Conflict occurred in 85.83% of the cases in the medium and 88.57% of the cases in the high segregation condition (from unweighted outcomes; $t=0.613$, two-tailed $p=0.541$).

Contribution rates by segregation conditions are summarized in Table 4. The differences between segregation conditions are the result of internalized *and* external social control. In order to test whether internalized social control can alone cause such differences between segregation conditions, results from Parts I and II are compared. The comparison reveals that minimal contact made an increase in contribution rates. The difference is significant at the 5% level, but not at the 1% level ($t=1.722$, one-tailed $p=0.043$). In Part II, the contribution rate was highest in the medium segregation condition, which *contradicts the segregation hypothesis*. Table 4 also shows average contribution rates in Parts III and IV of the experiment. The hypothesis that contribution rates are the same in the different conditions can be rejected both in Part III

Table 4. Average Contribution Rates in Different Segregation Conditions and Parts of the Experiment

incentives introduced first	segregation level low	medium	high	Total
*Part I**	49.64% (280)	51.81% (386)	46.61% (339)	49.45% (1005)
Part II	50.35% (282)	55.84% (385)	52.84% (335)	53.29% (1002)
Part III				
b (confirmation)	-	58.42% (190)	47.33% (150)	53.53% (340)
s (sel. incentives)	-	63.82% (199)	75.66% (189)	69.59% (388)
Part III total	40.35% (285)	61.18% (389)	63.13% (339)	55.97% (1013)
Part IV				
b (confirmation)	-	62.63% (190)	68.00% (150)	65.00% (340)
s (sel. incentives)	-	71.00% (200)	81.48% (189)	76.09% (389)
Part IV total	25.96% (285)	66.92% (390)	75.52% (339)	58.28% (1014)
Total (without Part I)	38.85% (852)	61.34% (1164)	63.87% (1013)	55.86% (3029)
Total	41.52% (1132)	58.97% (1550)	59.54% (1352)	54.26% (4034)

Notes. The number of cell-relevant cases is in parentheses. All human decisions are included.
* In Part I, subjects did not know their group membership and they did not see each other. Therefore their partition into the different segregation conditions only illustrates coincidental baseline contribution rates in the different experimental sessions.

(ANOVA F (2, 1010)=30.800, p<0.001) and in Part IV (ANOVA F (2, 1011)=108.721, p<0.001). It was predicted that the introduction of monetary selective incentives would result in higher contribution rates than when behavioral confirmation is introduced in Part III. Results confirm this hypothesis (t=4.487, one-tailed p<0.001). Furthermore, earlier introduction of monetary in-group selective incentives made a difference also in Part IV (t=3.285, two-tailed p=0.001). This result indicates that history effects still play a role in determining individual decision, despite the lack of feedback regarding the results of single-shot games. Furthermore, figures in Table 4 *also support the hypothesis that in the presence of monetary in-group selective incentives, the effect of segregation is stronger than in the presence of monetary behavioral confirmation*. In Part III, in the monetary in-group selective incentives condition average contribution rates are higher in the high segregation condition (75.66%) than in medium segregation (63.82%). On the other hand, in the monetary behavioral confirmation condition average contribution rates are higher in the medium segregation condition (58.42% vs. 47.33%).

Table 5. Results of Multilevel Logistic Regression on Contribution Propensities

independent variable	hypothesis about the direction of effect	multilevel model with fixed slopes of main effects	multilevel model assuming random slopes of social control effects
FIXED EFFECTS			
α baseline contribution propensity	?	-.038 (.082)	-.037 (.082)
s_0 internalized selective incentives	+	.109 (.072)	.117 (.072)
s external selective incentives	+	.407*** (.088)	.363*** (.104)
b_0 internalized behavioral confirmation	+	.617*** (.065)	.640*** (.077)
b external behavioral confirmation	+	.619*** (.104)	.615*** (.118)
t_0 internalized traitor rewards	-	-.175** (.055)	-.173** (.057)
RANDOM EFFECTS			
inter-individual variance σ^2		.616+++ (.085)	.628+++ (.121)
σ^2_{ui} (s_0)			.000 (.000)
σ^2_{ui} (s)			.300++ (.139)
σ^2_{ui} (b_0)			.196+++ (.093)
σ^2_{ui} (b)			.326+++ (.226)
σ^2_{ui} (t_0)			.009 (.050)
Covariances are reported below			
-2 Log Likelihood model		4480	4430
Improvement χ^2 (df in parentheses)		939*** (5)#	50*** (20)

Table 5b. Random Effects: Estimated Covariances

σ_{uxy}	ε_i	s_0	s	b_0	b
s_0	.000 (.000)				
s	-.252 (.108)	.000 (.000)			
b_0	.147 (.083)	.000 (.000)	-.194 (.085)		
b	-.359++ (.131)	.000 (.000)	.128 (.132)	-.079 (.116)	
t_0	-.005 (.072)	.000 (.000)	.425 (.153)	-.169 (.109)	.176 (.165)

Notes. N=4011 decisions for 203 subjects. Iterative Generalized Least Squares estimates. Numbers in parentheses are standard errors. ** significant at the 1% level, *** significant at the 0.1% level (two-tailed).
For testing random effects deviance tests are used: ++ significant at the 1% level, +++ significant at the 0.1% level (significance of difference in deviance compared to model without random slopes, for random covariates deviance is compared to model without random covariates).
#Basis of comparison: baseline multilevel logistic regression expressed in equation (2); α: 0.174** (0.066); σ2: 0.674+++ (0.087).

Analysis of contribution propensities: a simple model

To understand the underlying mechanisms of the segregation effect on intergroup conflict, individual decisions have to be analyzed. The first model in Table 5 reports results for the two-level model on contribution propensities without control variables.[9] The second model assumes that estimates of social control over subjects are normally distributed around their mean. In this model the variances and covariances are estimated as random effects. All human decisions except 23 cases (0.006%) are included. In these 23 cases subjects did not present any expectations about the behavior of their neighbors. In total, 4011 decisions are included in the analysis for 203 subjects.

The two models provide similar estimates. All effects are in the predicted direction. Hypotheses about the existence of internalized behavioral confirmation and internalized out-group selective incentives are supported. This means that contribution rates have increased with the difference between the number of expected in-group contributors and defectors and they have decreased with the number of out-group contacts. The effect of internalized in-group selective incentives is not significant. According to this result, the number of in-group contacts does not enforce contributions, if one controls for internalized behavioral confirmation. As predicted, both forms of external social control have a significant effect. It is important to note, however, that this simple model did not include any control variables.

Contribution rates between subjects have a high unexplained variance.[10] The influence of behavioral confirmation and monetary in-group selective incentives varies significantly between subjects. The hypothesis that the sizes of internalized selective incentives are the same for the subjects cannot be rejected. High positive deviations from the average baseline contribution rate are correlated with negative deviations from the average importance of monetary rewards for confirmation. This is not surprising because subjects, who evaluate monetary gains less, contribute more to the success of their group.

The effect of personal characteristics and other control variables

To see which personal characteristics are responsible for high inter-individual variation, the model is extended by background variables and certain attitude measures. Furthermore, in the previous analysis it was assumed that intra-individual variation (ξ_{ir}) has a zero expected value and it is independent from the decision round r. If contribution propensities are not stable in the single-shot games within experimental parts, then an independent trend element has to be included in the analysis and the assumption that intra-individual variation (ξ_{ir}) has a zero expected value has to be relaxed. As parts were separated by breaks, instead of checking for a single learning trend, it is better to distinguish between a within part and a between part learning trend in the analysis.

Two analyses are conducted again: one assuming fixed social control effects without random variation and another assuming a random variation and covariation of these estimates (see Table 6). As the analysis controls for some disturbing procedural effects, results show the net effect of main variables.

There are remarkable changes in the parameter estimates of social control. The effect of internalized in-group selective incentives became significant and the significant effect of internalized out-group selective incentives has disappeared. The large increase in the estimate of baseline contribution propensity (constant) also indicates that the omission of independent trends resulted in a systematic bias in previous estimates in Table 5. Because of the negative between parts tendency, the baseline contribution rate was underestimated and the decrease between Part I and Part II was attributed to the effect of internalized out-group selective incentives. On the basis of the analysis reported in Table 6, after controlling for a negative learning tendency, it turns out that on average, out-group selective incentives in an internalized form do not influence the decision of subjects. On the other hand, this interpretation and also the confirmation of the existence of internalized in-group selective incentives has to be handled with reservations. The inclusion of a between parts trend in a linear functional form in the analysis does not stand on a firm theoretical basis. Furthermore, since the high correlation with experimental manipulations (the introduction of minimal contact and monetary forms of social control), the learning effect might include part of influence that should be attributed to other variables.

There is another complication in relation to the difference in contribution propensities between Parts I and II. Silent identification (Bohnet and Frey, 1999) enters social dilemma experiments, when subjects are able to see each other. The visibility of others decreases social distance, allows for empathy and helps to conceptualize the experimental situation. However, this effect cannot be separated from the influence of internalized selective incentives that are not contingent on predictions. If silent identification is a valid mechanism in the IPG game, the analysis overestimates the effect of internalized selective incentives. The unexpected positive sign of the t_0 estimate can also partly be explained by silent identification.

Among personal background variables, gender has no significant effect, although simple descriptive statistics showed that women had higher contribution rates (55.94%) than men (52.14%). Based also on descriptive statistics, subjects who already graduated were more contributive (61.54%) than students (53.58%). This effect is not significant in the model, as it is ruled out by other variables, mainly by social orientat-

[9] For goodness-of-fit, -2 log likelihood statistics and $\chi 2$ tests of improvement are indicated at the bottom of tables.

[10] For testing hypotheses about random effects it is more appropriate to use deviance tests than the t-test (cf. van Duijn, van Busschbach, and Snijders, 1999: 192-193).

Table 6a. Results of Multilevel Logistic Regression on Contribution Propensities with Personal Characteristics and Procedure Effects

independent variable	hypothesis about the direction of effect	multilevel model with fixed slopes of main effects		multilevel model random slopes of main effects	
FIXED EFFECTS					
α (constant) baseline contr. propensity	?	1.378^{**}	(.423)	1.516^{***}	(.409)
s_0 internalized in-group selective incentives	+	$.186^{*}$	(.082)	$.188^{*}$	(.081)
t_0 internalized out-group selective incentives	-	.165	(.086)	.142	(.086)
b_0 internalized behavioral confirmation	+	$.586^{***}$	(.067)	$.591^{***}$	(.080)
s monetary in-group selective incentives	+	$.769^{***}$	(.109)	$.699^{***}$	(.127)
b monetary behavioral confirmation	+	$.718^{***}$	(.108)	$.705^{***}$	(.126)
Personal characteristics and other subject-level variables					
gender (1=male)		-.176	(.143)	-.196	(.137)
student at the university (1=yes)		-.219	(.370)	-.352	(.357)
studies at the law faculty		-.109	(.366)	-.015	(.351)
studies natural sciences		-.057	(.344)	-.065	(.330)
studies economic, business, or spatial sci.		-.030	(.335)	.095	(.322)
studies social sciences		.068	(.309)	.136	(.296)
student of literary studies or arts		.056	(.316)	.133	(.303)
did a similar experiment before		-.154	(.136)	-.188	(.131)
strong risk aversion towards gains		-.163	(.135)	-.180	(.129)
strong loss aversion		.115	(.134)	.132	(.128)
consistent answers on social orientation qs		$-.374^{*}$	(.181)	$-.400^{*}$	(.173)
prosocial orientation		$.511^{**}$	(.183)	$.487^{**}$	(.175)
egalitarian orientation		$.388^{*}$	(.176)	$.392^{*}$	(.169)
number of acquainted subjects in the exp.		-.079	(.088)	-.093	(.085)
delay (minutes) at the start of the exp.		.008	(.007)	.006	(.007)
quiz questions answered correctly %		-.005	(.004)	-.005	(.004)
Procedure effects					
within part trend		$-.215^{***}$	(.036)	$-.213^{***}$	(.036)
endgame effect		$.373^{**}$	(.125)	$.370^{**}$	(.126)
between parts trend		$-.397^{***}$	(.060)	$-.379^{***}$	(.061)
last iterated game was a draw		$.538^{***}$	(.149)	$.515^{***}$	(.152)
last iterated game was lost		.185	(.122)	.199	(.125)
last iterated game was won		.214	(.123)	$.275^{*}$	(.125)
RANDOM EFFECTS					
inter-individual variance σ^2		$.574^{+++}$	(.083)	$.559^{+++}$	(.116)
σ^2 *ui* (s_0)				.000	(.000)
σ^2 *ui* (t_0)				.002	(.050)
σ^2 *ui* (b_0)				$.202^{+++}$	(.096)
σ^2 *ui* (*s*)				$.322^{+++}$	(.152)
σ^2 *ui* (*b*)				$.421^{+++}$	(.246)
Covariances are reported below					+
-2 Log Likelihood model		4480			4430
Improvement χ^2 (df in parentheses)		$939^{***}(5)^{\#}$			$50^{***}(20)$

Table 6b. Random Effects: Estimated Covariances

σ_{uxy}	ε_i	s_0	t_0	b_0	s
s_0	.000 (.000)				
t_0	-.018 (.071)	.000 (.000)			
b_0	.037 (.083)	.000 (.000)	-.054 (.117)		
s	-.163 (.109)	.000 (.000)	.476 (.169)	$-.192^{+}$ (.090)	
b	$-.287^{+}$ (.133)	.000 (.000)	.152 (.180)	-.084 (.123)	.063 (.143)

Notes. N=4011 decisions for 203 subjects. Iterative Generalized Least Squares estimates. Numbers in parentheses are standard errors.
* significant at the 5% level, ** significant at the 1% level, *** significant at the 0.1% level (two-tailed).
For testing random effects deviance tests are used: $^{+}$ significant at the 5% level, $^{+++}$ significant at the 0.1% level (significance of difference in deviance compared to model without random slopes, for random covariates deviance is compared to model without random covariates)

tions. The analysis of college major does not reveal an effect of economics training. The argument that experience matters at all is questioned by the insignificant effect of participating in a similar experiment before. Again, the difference in descriptive statistics (56.14% vs. 51.44%) could be explained by selection on attitude measures.

Subjects were characterized as strongly risk-averse, if they chose for risk-averse alternatives both in simple and complex gambles. 91 subjects (45.3%) were strongly risk-averse towards gains, 92 (45.8%) were strongly risk-averse towards mixed gambles, and 83 (39.5%) were strongly risk-seeking towards losses. Effects of risk-aversion and loss-aversion, however, are not significant in the models.

The only personal characteristics that are found significant in explaining contribution propensities are social orientations. For questions about social orientations, 77 (37.9%) subjects gave inconsistent answers. Inconsistency was a significant predictor of contribution rates, which is probably related to the relevance of calculation abilities. Among subjects, who gave consistent answers, 76 (61.3%) were prosocial, which is higher than in previous experiments (for an overview see Schulz and May, 1989). As an exception, Liebrand (1984) found a similar high rate in his experiments conducted in Groningen. Results clearly support the argument that prosocial (and also egalitarian) orientation is primarily directed towards in-group members and therefore increases contribution rates in the IPG game. The strong effects also indicate that social orientations are important predictors of behavior in intergroup situations. Individuals with prosocial and egalitarian attitudes seem to be responsible for the emergence of mutually harmful outcomes.

There was no significant effect of delay time at the start of the experiment and of how many others were acquainted to subjects in the laboratory. These factors that are related to the experimental environment did not disturb the behavior of subjects.

Although Bayesian learning effects cannot enter the series of single-shot games, as experimental time passes, subjects might understand the structure of the game better and can become more experienced with the decision task. Previous experiments of iterated PD, public good, and IPG games found that subjects approach the all-defection equilibrium over time (Isaac, McCue, and Plott, 1985; Andreoni, 1988; Andreoni and Miller, 1993; Bornstein, Winter, and Goren, 1996; Goren and Bornstein, 2000; Goren, 2001), which results in decreasing cooperation rates. In this study, a decay of contribution is found for the series of single-shot games. Contribution rates decreased for those, who had some misunderstanding of the task before the game, but also for those, who answered quiz questions correctly. Besides the decreasing within part trend, in the last round of every part contribution rates increased significantly. This is a surprising result, since subjects knew that the outcome of the last round would not be announced. This is exactly the opposite of what would be predicted on the basis of arguments of traditional game theory even if subjects had the incorrect perception that they are playing repeated games. By analyzing last rounds only, model parameters were similar to those values that were reported in Table 6, including an insignificant effect of internalized in-group selective incentives. It means that higher contribution propensities in the last rounds cannot be explained by the reduction of cognitive dissonance ("in the last round I have to be nice, otherwise I cannot look at my fellow neighbors"). The resulting U-shape trend, however, has some correspondence to experimental findings in the iterated two-person PD and in collective action games (Rapoport and Chammah, 1965; Guttman, 1986).

Besides a within part trend, a between parts trend is also included in the models in Table 6 as a control variable. Both trends are highly significant, as well as the puzzling endgame effect. Trends and endgame effects are not the only unexpected procedure effects. After controlling for the results of repeated games, it emerged that a mutually harmful draw (punishment) "burns in" the memory of subjects and increases contribution propensities also in the single-shot games. Unfortunately, this points to a weakness of the present design. This also indicates that subjects use their long-term memory to estimate whether or not their decision could make a difference for the outcome in the forthcoming single-shot game. If they believe that a draw will occur, a single individual contribution can turn the outcome to winning the public good.

Interaction effects

As Table 6 demonstrated, the significant effect of internalized out-group selective incentives disappeared after the inclusion of learning trends. It might be possible that this form of social control is mistakenly conceptualized and out-group selective incentives have a different nature. They might stem from the presence of the other group as a whole or they exist only in certain dyadic relations.

The extension of the model by interaction effects helps with some clarification (see Table 7). It seems that internalized out-group selective incentives are activated in the dyadic context, but not in every neighborhood relation. Only neighbors of the opposite sex provide a significant control in the form of out-group selective incentives. This indicates that internalized pressure against contribution in the presence of opposite group members is activated only, when a *substantive distinction can be made apart from minimal group membership*. Gender is possibly the most apparent characteristic that can be the source of this distinction between strangers. With respect to the interaction between gender and internalized behavioral confirmation, no significant effect is found on contribution propensities.

However, descriptive statistics showed that subjects expected contribution more from in-group neighbors of the same sex and additionally, women were expected to contribute more.

Acquainted neighbors did not experience stronger social control than unknown ones did. Similar to the insignificant effect of the number of acquainted subjects in the experiment,

Table 7a. Results of Multilevel Logistic Regression on Contribution Propensities with Personal Characteristics, Procedure Effects, and Cross-level Interactions

independent variable	hypothesis about the direction of effect	multilevel model with fixed slopes		multilevel model random slopes	
FIXED EFFECTS					
Main variables					
α (constant) baseline contr. propensity	?	1.346***	(.402)	1.491**	(.477)
s_0 internalized in-group selective incentives	+	.176*	(.082)	.165*	(.084)
t_0 internalized out-group selective incentives	-	.223	(.132)	.238	(.134)
b_0 internalized behavioral confirmation	+	.589***	(.119)	.618***	(.141)
s monetary in-group selective incentives	+	.769***	(.110)	.745***	(.135)
b monetary behavioral confirmation	+	.703***	(.109)	.681***	(.125)
Personal characteristics and other subject-level variables					
gender (1=male)		-.089	(.146)	-.135	(.143)
student at the university (1=yes)		-.177	(.372)	-.201	(.364)
studies at the law faculty		-.162	(.368)	-.136	(.360)
studies natural sciences		-.101	(.349)	-.161	(.341)
studies economic, business, or spatial sci.		-.080	(.339)	-.002	(.330)
studies social sciences		-.001	(.312)	.000	(.305)
student of literary studies or arts		.045	(.317)	.066	(.309)
did a similar experiment before		-.179	(.136)	-.221	(.133)
strong risk aversion towards gains		-.172	(.134)	-.157	(.132)
strong loss aversion		.131	(.133)	.164	(.131)
consistent answers on social orientation qs		-.397*	(.180)	-.404*	(.176)
prosocial orientation		.330	(.206)	.353	(.202)
egalitarian orientation		.419*	(.203)	.394*	(.200)
number of acquainted subjects in the exp.		-.066	(.089)	-.066	(.087)
delay (minutes) at the start of the exp.		.006	(.007)	.006	(.007)
quiz questions answered correctly %		-.004	(.005)	-.005	(.005)
Procedure effects					
within part trend		-.178	(.121)	-.188	(.122)
endgame effect		.379**	(.126)	.381**	(.127)
between parts trend		-.397***	(.061)	-.386***	(.062)
last iterated game was a draw		.527***	(.150)	.495**	(.157)
last iterated game was lost		.180	(.123)	.186	(.128)
last iterated game was won		.214	(.124)	.266*	(.128)
Cross-level interactions					
t_0 × number of acquainted opposite neighbors		-.153	(.196)	-.164	(.194)
b_0 × number of acquainted in-group neighbors		.302	(.261)	.338	(.312)
t_0 × number of opposite neighbors of the other sex		-.351**	(.134)	-.373**	(.137)
t_0 × number of male opposite neighbors		.191	(.134)	.156	(.136)
b_0 × number of in-group neighbors of the same sex		-.038	(.084)	-.128	(.102)
b_0 × number of female in-group neighbors		.302	(.261)	.017	(.108)
t_0 × prosocial orientation		.275*	(.131)	.256*	(.132)
b_0 × prosocial orientation		.052	(.134)	.098	(.161)
t_0 × egalitarian orientation		-.057	(.149)	-.025	(.149)
b_0 × egalitarian orientation		.039	(.143)	.004	(.172)
within part trend ´ quiz questions correct %		.000	(.001)	.000	(.001)
RANDOM EFFECTS					
inter-individual variance σ^2		.563^{+++}	(.082)	.512^{+++}	(.084)
σ^2 *ui* (s_0)				.000	(.000)
σ^2 *ui* (t_0)				.000	(.000)
σ^2 *ui* (b_0)				.143^{+++}	(.089)
σ^2 *ui* (*s*)				.549^{+++}	(.187)
σ^2 *ui* (*b*)				.379^{+++}	(.240)
Covariances are reported below				+	
-2 Log Likelihood model		4211		4169	
Improvement χ^2 (df in parentheses)				42**	(20)
vs. previous model		36***	(11)	29**	(11)

Table 7b. Random Effects: Estimated Covariances

σ_{uxy}	ε_i	s_0	t_0	b_0	s
s_0	.000 (.000)				
t_0	.000 (.000)	.000 (.000)			
b_0	.004 (.072)	.000 (.000)	.000 (.000)		
s	.037 (.107)	.000 (.000)	.000 (.000)	-.145 (.093)	
b	-.200^{++} (.118)	.000 (.000)	.000 (.000)	-.031 (.116)	.201 (.152)

Notes. N=4011 decisions for 203 subjects. Iterative Generalized Least Squares estimates. Numbers in parentheses are standard errors.
* significant at the 5% level, ** significant at the 1% level, *** significant at the 0.1% level (two-tailed).
For testing random effects deviance tests are used: $^{+}$ significant at the 5% level, $^{+++}$ significant at the 0.1% level (significance of difference in deviance compared to model without random slopes, for random covariates deviance is compared to model without random covariates)

this result can probably be attributed to the fact that they were not close acquaintances or simply, subjects considered laboratory conditions impersonal. Prosocial and egalitarian attitudes were not correlated with higher relative weight of internalized social control. Only the interaction between out-group selective incentives and prosocial orientation proved to be significant. This effect indicates that prosocial subjects liked to be "local heroes", who contributed even when they were surrounded by members of the other group. This is another indication of how prosocial attitudes can be harmful in the intergroup context.

DISCUSSION

The main objective of this study was to show how social control mechanisms enter into simple experimental situations and can affect individual decisions in social dilemmas. As an aggregated result of different forms of social control, it was demonstrated *how network segregation might induce the emergence of conflict between groups*. To discover the underlying mechanisms, the study investigated what is the exact nature of *social control* and what are the forms that are already present in a condition with only minimal contact between subjects. For the test of hypotheses, a unique experimental design was introduced based on special arrangements in the laboratory. With this setup, network based social control, which is believed to be influential also in real life, was the target of analysis in an experimental environment.

In the experiments, intergroup competitions were modeled by an Intergroup Public Goods game (Rapoport and Bornstein, 1987; Takács, 2001). Comparison of segregation conditions showed that intergroup conflict was least likely in a completely mixed setting and was most likely when members of the groups were arranged according to a segregated pattern, which confirms the segregation hypothesis. Furthermore, as predicted, the segregation effect was stronger in the presence of monetary in-group selective incentives than in the presence of monetary behavioral confirmation.

By analyzing individual decisions, social control mechanisms were uncovered that cause the segregation effect on the aggregated level. *Behavioral confirmation* is found to be the form of social control, which strongly affects individual contribution propensities, also in an internalized form. Subjects adjusted their decisions towards the expected decision of their in-group contacts even when only a minimal contact and "minimum network relations" have been established between them. Estimates of model parameters indicate that under the chosen reward structure, internalized behavioral confirmation affected contribution propensities as much as monetary behavioral confirmation. Concerning behavioral confirmation, however, a part of the significant effect might be due to the bi-directional relationship between own behavior and expectations about the behavior of others. Subjects formulated their expectations at the same time of their decisions; therefore the guess what others do is not obviously an exogenous variable. Subjects, for instance, could have formulated their expectations in order to avoid cognitive dissonance or to project their decision on others. This might have played a role for some subjects,[11] but it sounds unlikely that many subjects fitted their expectations to their behavior, which does not pay off, and not the behavior to expectations, which does.

Besides, no strong support was found for the presence of other forms of internalized social control. Internalized in-group selective incentives had a significant effect after controlling for a between parts trend. Internalized out-group selective incentives might be activated in a dyad with minimal contact, but it is not a general mechanism. Its clear presence was found only between neighbors of the opposite sex. External social control that was introduced in a form of additional monetary incentives had a significant effect.

Contribution rates in the minimal contact condition were highest in the medium segregation condition, which is a somewhat puzzling result. A possible explanation is that there is a *ceiling effect*, which means that a presence of a single in-group neighbor activates sufficient internalized social control to enhance contribution to almost full certainty. This explanation is supported by evidence of high likelihood of conflict in the medium segregation condition (cf. Table 3). Another reason might be that the strength of internalized social control is a nonlinear function of the number of in-group contacts. As a consequence, there is a marginal decrease in the segregation

[11] Only one subject revealed such motivations in the post-experiment questionnaire.

effect on the likelihood of intergroup conflict and medium levels of segregation are already associated with harmful outcomes.

Among personal characteristics, only social orientations had significant effects. Subjects with prosocial and egalitarian attitudes were more contributive and consequently were also more responsible for the emergence of mutually harmful outcomes between the groups than others. Another indication of that prosocial orientations are correlated with more generous behavior for the in-group, but more hostile behavior towards the out-group, is the positive interaction effect of out-group selective incentives and prosocial orientation. This implies that subjects with prosocial orientation behave more likely as local heroes. If members of the other group surround them, they do not surrender at all. As a macro consequence, mutually harmful outcomes can occur even in the case of complete mixing, if there are enough prosocial individuals.

To summarize, the present study demonstrated that laboratory experiments with minimal contact between subjects provide an important insight for understanding network effects and the influence of internalized social control in intergroup situations. Results support policy arguments to promote interethnic relations and decrease segregation in order to help conflict resolution.

REFERENCES

Allport GW. 1954. *The Nature of Prejudice*. Reading, MA: Addison-Wesley.

Andreoni J. 1988. Why free ride? Strategies and learning in public goods experiments. *Journal of Public Economics* 37: 291-304.

Andreoni J & Miller JH. 1993. Rational cooperation in the finitely repeated prisoner's dilemma: experimental evidence. *Economic Journal* 103: 570-585.

Asch S. 1956. *Studies of independence and conformity*. Washington, DC: American Psychological Ass'n.

Black DJ. 1984. Social Control as a Dependent Variable. In: Black, DJ. (Ed.): *Towards a General Theory of Social Control*. vol.1. New York, Academic.

Bohnet I & Frey BS. 1999. The sound of silence in prisoner's dilemma and dictator games. *Journal of Economic Behavior and Organization* 38: 43-57.

Bornstein G & Ben-Yossef M. 1994. Cooperation in intergroup and single-group social dilemmas. *Journal of Experimental Social Psychology* 30: 52-67.

Bornstein G, Erev I, & Rosen O. 1990. Intergroup competition as a structural solution to social dilemmas. *Social Behavior* 5: 247-260.

Bornstein G, Winter E, & Goren H. 1996. Experimental study of repeated team-games. *European Journal of Political Economy* 12: 629-639.

Bryk AS & Raudenbush SW. 1992. *Hierarchical linear models: Applications and data analysis methods*. London, Sage.

Budescu DV, Rapoport Am, & Suleiman R. 1990. Resource dilemmas with environmental uncertainty and asymmetric players. *European Journal of Social Psychology* 20: 475-487.

Burt RS. 1998. The gender of social capital. *Rationality and Society* 10: 5-46.

Carter J & Irons M. 1991. Are economists different, and if so, why? *Journal of Economic Perspectives* 5: 171-177.

Chong D. 1991. *Collective action and the civil rights movement*. Chicago, Chicago University Press. Chwe MS-Y. 1999. Structure and strategy in collective action. *American Journal of Sociology* 105: 128-156.

Dawkins, R. 1976. *The selfish gene*. Oxford, Oxford University Press. Doi, T. 1994. Social orientation analysis of the common and individual interest problem. In Schulz U, et al. (Eds.), *Social dilemmas and cooperation*. Berlin, Springer Verlag.

Drolet AL & Morris MW. 2000. Rapport in conflict resolution: accounting for how face-to-face contact fosters mutual cooperation in mixed-motive conflicts. *Journal of Experimental Social Psychology* 36(1): 26-50.

Eckel CC & Grossman PJ. 1998. Are women less selfish than men? Evidence from dictator experiments. *Economic Journal* 108: 726-735.

Farquhar PH. 1984. Utility assessment methods. *Management Science* 30: 1283-1300.

Finkel SE & Opp K-D. 1991. Party identification and participation in collective political action. *Journal of Politics* 53: 339-371.

Flache A & Macy MW. 1996. The weakness of strong ties: Collective action failure in a highly cohesive group. *Journal of Mathematical Sociology* 21: 3-28.

Frank RH, Gilovich T, & Regan D. 1993. Does studying economics inhibit cooperation? *Journal of Economic Perspectives* 7: 159-171.

Gächter S & Fehr E. 1999. Collective action as a social exchange. *Journal of Economic Behavior and Organization* 39: 341-369.

Gibbs JP. 1981. *Norms, Deviance, and Social Control: Conceptual Matters*. New York, Elsevier.

Goldstein H. 1995. *Multilevel statistical models*. London, Edward Arnold Publishers.

Goren H. 2001. The effect of out-group competition on individual behavior and out-group perception in the Intergroup Prisoner's Dilemma (IPD) game. *Group Processes and Intergroup Relations* 4: 160-182.

Goren H & Bornstein G. 2000. The effects of intragroup communication on intergroup cooperation in the repeated Intergroup Prisoner's Dilemma (IPD) game. *Journal of Conflict Resolution* 44: 700-719.

Gould RV. 1993. Collective action and network structure. *American Sociological Review*, 58: 182-196.

Gould RV. 1995. *Insurgent identities: Class, community, and protest in Paris from 1848 to the Commune*. Chicago, University of Chicago Press.

Größer J. & Schram, A. 2006. Neighborhood Information Exchange and Voter Participation: An Experimental Study. *American Political Science Review*, forthcoming.

Guttman JM. 1986. Matching behavior and collective action: Some experimental evidence. *Journal of Economic Behavior and Organization* 7: 171-198.

Heckathorn DD. 1990. Collective sanctions and compliance norms: A formal theory of group-mediated social control. *American Sociological Review* 55: 366-384.

Heckathorn DD. 1993. Collective action and group heterogeneity: Voluntary provision versus selective incentives. *American Sociological Review* 58: 329-350.

Insko CA, Schopler J, Graetz KA, Drigotas SM, Currey DP, Smith SL, Brazil D, & Bornstein G. 1994. Interindividual-intergroup discontinuity in the prisoner's dilemma game. *Journal of Conflict Resolution* 38: 87-116.

Isaac RM, McCue KF, & Plott C. 1985. Public goods provision in an experimental environment. *Journal of Public Economics* 26: 51-74.

Kornhauser RR. 1978. *Social Sources of Delinquency: An Appraisal of Analytic Models*. Chicago, University of Chicago Press.

Kosfeld M. 2003. *Network Experiments*. Institute for Empirical Research in Economics, University of Zürich Working Paper Series, No. 152.

Kuran T. 1995. *Private truths, public lies: The social consequences of preference falsification*. Cambridge, MA: Harvard University Press.

Ledyard JO. 1995. Public goods: A survey of experimental research. In Kagel JH & Roth A E. (Eds.), *The handbook of experimental economics*. Princeton: Princeton University Press.

Liebrand WBG. 1984. The effect of social motives, communication and group size on behavior in an N-person multi-stage mixed-motive game. *European Journal of Social Psychology* 14: 239-264.

Lovaglia MJ, Willer R, & Troyer L. 2003. Power, Status, and Collective Action: Developing Fundamental Theories to Address a Substantive Problem. In Thye SR, Skvoretz J, & Lawler EJ. (Eds.), *Advances in Group Processes*, 20. Greenwich, CT: JAI Press.

Macy MW. 1993. Backward looking social control. *American Sociological Review* 58: 819-836.

Marwell G & Ames R. 1981. Economists free ride, does anyone else? Experiments on the provision of public goods, IV. *Journal of Public Economics* 15: 295-310.

Marwell G & Oliver PE. 1993. *Critical mass in collective action*. Cambridge: Cambridge University Press.

Marwell G, Oliver PE, & Prahl R. 1988. Social networks and collective action: A theory of the critical mass. III. *American Journal of Sociology* 94: 502-534.

McAdam D. 1986. Recruitment to high-risk activism: The case of freedom summer. *American Journal of Sociology* 92: 64-90.

McCusker C & Carnevale PJ. 1995. Framing in Resource Dilemmas: Loss Aversion and the Moderating Effects of Sanctions. *Organizational Behavior and Human Decision Processes* 61(2): 190-201.

Nowell C. & Tinkler S. 1994. The influence of gender on the provision of a public good. *Journal of Economic Behavior and Organization* 25: 25-36.

Oberschall AR. 1994. Rational choice in collective protests. *Rationality and Society* 6: 79-100.

Opp K-D. 1989. *The rationality of political protest: A comparative analysis of rational choice theory*. Boulder, CO: Westwiew.

Ortmann A & Tichy LK. 1999. Gender differences in the laboratory: Evidence from prisoner's dilemma games. *Journal of Economic Behavior and Organization* 39: 327-339.

Pingle M. 1995. Imitation versus rationality: An experimental perspective on decision making. *Journal of Socio-Economics* 24: 281-316.

Rapoport Am & Bornstein G. 1987. Intergroup competition for the provision of binary public goods. *Psychological Review* 94: 291-299.

Rapoport Am, Bornstein G, & Erev I. 1989. Intergroup competition for public goods: Effects of unequal resources and relative group size. *Journal of Personality and Social Psychology* 56: 748-756.

Rapoport An & Chammah AM. 1965. *Prisoner's dilemma: A study in conflict and cooperation*. Ann Arbor, MI: University of Michigan Press.

Raub, W & Snijders C. 1997. Gains, losses, and cooperation in social dilemmas and collective action: The effects of risk preferences. *Journal of Mathematical Sociology* 22: 263-302.

Rege M & Telle K. 2004. The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics* 88: 1625-1644.

Sandell R & Stern C. 1998. Group size and the logic of collective action: A network analysis of a Swedish temperance movement 1896-1937. *Rationality and Society* 10: 327-345.

Schopler J & Insko CA. 1992. The discontinuity effect in interpersonal and intergroup relations: Generality and mediation. *European Review of Social Psychology* 3: 121-151.

Schulz U & May T. 1989. The recording of social orientations with ranking and pair comparison procedures. *European Journal of Social Psychology* 19: 41-59.

Scott JF. 1971. *Internalization of norms*. Englewood Cliffs, NJ: Prentice-Hall.

Solnick SJ & Schweitzer ME. 1999. The influence of physical attractiveness and gender on ultimatum game decisions. *Organizational Behavior and Human Decision Processes* 79: 199-215.

Stockard J, van de Kragt AJ, & Dodge PJ. 1988. Gender roles and behavior in social dilemmas: Are there sex differences in cooperation and in its justification? *Social Psychology Quarterly* 51: 154-163.

Suleiman R & Or-Chen K. 1999. Providing step-level public goods under uncertainty: The case of probable external supply. In Foddy M, Smithson M, Schneider S, & Hogg M. (Eds.), *Resolving social dilemmas: Dynamic, structural, and intergroup aspects*. Philadelphia: Psychology Press / Taylor and Francis.

Takács K. 2001. Structural embeddedness and intergroup conflict. *Journal of Conflict Resolution* 45: 743-769.

Takács K. 2002. *Social networks and intergroup conflict*. Groningen, ICS Dissertation Series.

Tickle-Degnen L & Rosenthal R. 1990. The nature of rapport and its nonverbal correlates. *Psychological Inquiry* 1: 285-293.

van Assen MALM & Snijders C. 2004. Effects of risk preferences in social dilemmas: A game-theoretical analysis and evidence from two experiments. In: Suleiman R, Budescu D, Messick DM, & Fischer I. (Eds.): *Contemporary Approaches to Social Dilemma Research*. Cambridge, MA, Cambridge University Press.

van Duijn MAJ, van Busschbach JT, & Snijders TAB. 1999. Multilevel analysis of personal networks as dependent variables. *Social Networks* 21: 187-209.

van Lange PAM. 1999. The pursuit of joint outcomes and equality in outcomes: an integrative model of social value orientation. *Journal of Personality and Social Psychology* 77: 337-349.

van Lange PAM, Otten W, De Bruin EMN, & Joireman JA. 1997. Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology* 73: 733-746.

van de Kragt AJC, Dawes RM, & Orbell JM. 1988. Are people who cooperate 'rational altruists'? *Public Choice* 56: 233-247.

van de Kragt AJC, Orbell JM, & Dawes RM. 1983. The Minimal Contributing Set as a solution to public goods problems. *American Political Science Review* 77: 112-122.

Villareal A. 2002. Political competition and violence in Mexico: Hierarchical social control in local patronage structures. *American Sociological Review* 67: 477-498.

Walters AE, Stuhlmacher AF, & Meyer LL. 1998. Gender and negotiator competitiveness: A meta-analysis. *Organizational Behavior and Human Decision Processes* 76: 1-29.

Yamagishi T. 1986. The provision of a sanctioning system as a public good. *Journal of Personality and Social Psychology* 51(1): 110-116.