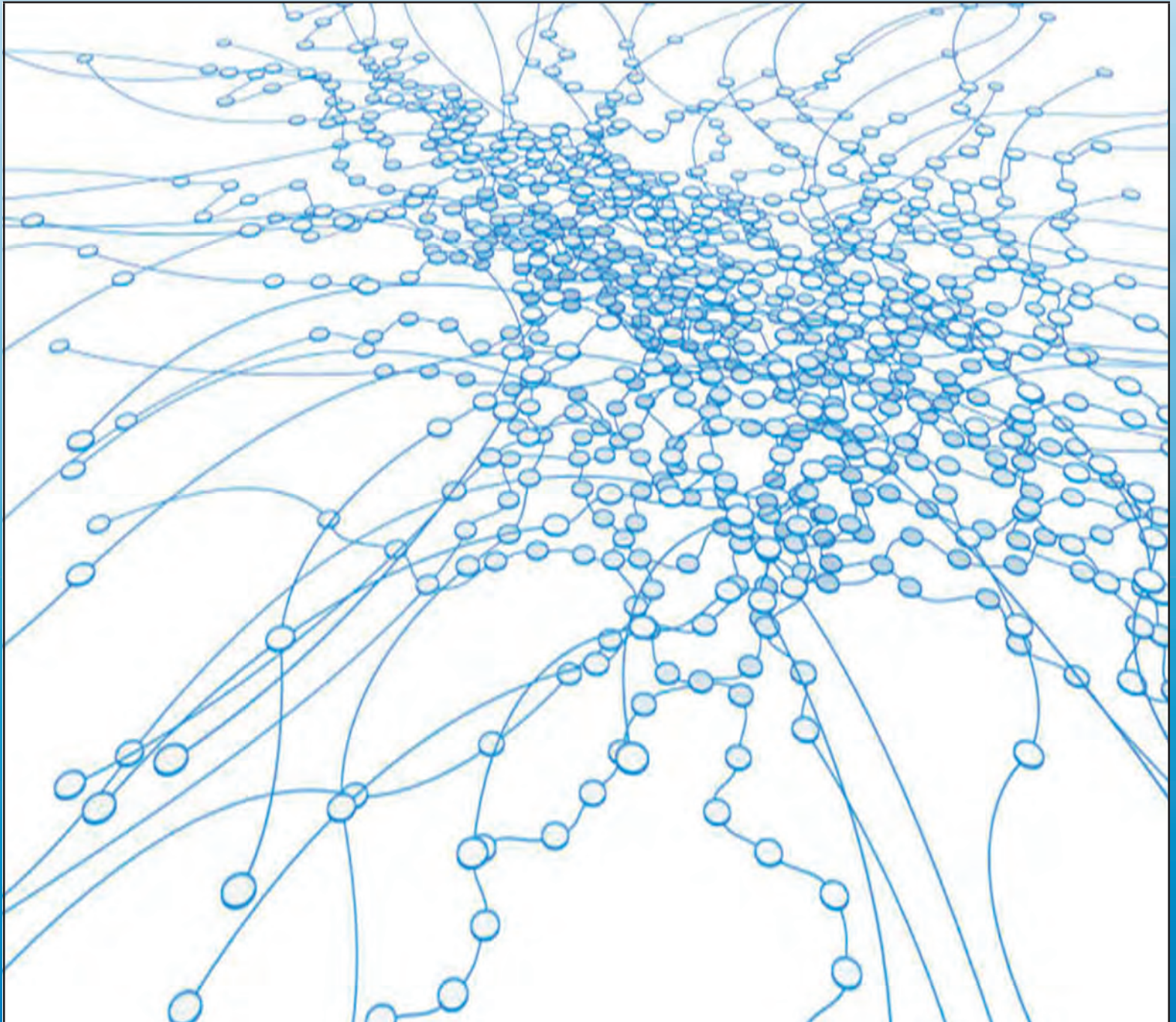


# Connections

Volume 33, Issue 1

July 2013



*Official Journal of the International Network for Social Network Analysts*

International Network for  
Social **Network** Analysis



**Connections** publishes original empirical, theoretical, and methodological articles, as well as critical reviews dealing with applications of social network analysis. The research spans many disciplines and domains including Communication, Anthropology, Sociology, Psychology, Organizational Behavior, Knowledge Management, Marketing, Social Psychology, Political Science, Public Health, Policy, Medicine, Physics, Economics, Mathematics, and Computer Science. As the official journal for the International Network for Social Network Analysis, the emphasis of the publication is to reflect the ever-growing and continually expanding community of scholars using network analytic techniques. **Connections** also provides an outlet for sharing news about social network concepts and techniques and new tools for research.



## Connections

Manuscripts selected for publication are done so based on a peer-review process. See Manuscript Submissions for instructions and formatting details. The journal is edited by the Connections Editorial Board:

**Dimitris Christopoulos, *Editor***

MODUL University, Vienna, Austria

**Kathryn Coronges, *Managing Editor***

United States Military Academy, West Point, NY

**Alyse Pulido, *Associate Editor***

United States Military Academy, West Point, NY

**Joseph Dunn, *Copy Editor***

ReEstablishment Co., Garrison, NY

***Editorial Headquarters***

MODUL University Vienna

Department of Public Governance & Management

Am Kahlenberg 1, 1190 Vienna, Austria

---

### Former Editors

**Barry Wellman, *Founding Editor***

Years of service: 1976-1988

**Susan Greenbaum, *Editor***

Years of service: 1987-1990

**Alvin Wolfe, *Editor***

Years of service: 1990-1993

**John Skvoretz, *Editor***

Years of service: 1993-1997

**Katherine Faust, *Editor***

Years of service: 1993-1997

**Steve Borgatti, *Editor***

Years of service: 1997-2000

**Bill Richards, *Editor***

Years of service: 2000-2003

**Tom Valente, *Editor***

Years of service: 2003-2011

### Editorial Board

**Bruce Cronin**

University of Greenwich

**Jana Diesner**

University of Illinois at UC

**Dan Halgin**

University of Kentucky

**Kayla de la Haye**

The Rand Corporation

**Nicholas Harrigan**

Singapore Management University

**Betina Hollstein**

University of Hamburg

**Karin Ingold**

University of Bern

**Susie J Pak**

St. John's University

**Juergen Pfeffer**

Carnegie Mellon University

**Ian McCulloh**

Curtin University

**Meredith Rolfe**

London School of Economics

**Kerstin Sailer**

University College London

**David Schaefer**

Arizona State University

**Christophe Sohn**

CEPS-Luxembourg

### DEN Section Editors

**Pacey Foster**

University of Massachusetts

**Rich de Jordy**

Northeastern University

---

**Connections** publishes original empirical, theoretical, tutorial, and methodological articles that use social network analysis. The journal publishes significant work from any domain that is relevant to social network applications and methods. Commentaries or short papers in response to previous articles published in the journal are considered for publication. Review articles that critically synthesize a body of published research are also considered, but normally are included by invitation only. Book reviews, network images, review articles, and letters to the editor are also welcome. Manuscripts selected for publication are done so based on a peer-review process.

# INSNA

International Network for Social Network Analysis

## **International Network for Social Network Analysis**

Connections is the official journal of the **International Network for Social Network Analysis (INSNA)**. INSNA is a scientific organization made up of scholars across the world. Updated information about INSNA's annual conference (**Sunbelt Social Network Conferences**) can be found on the website at [www.insna.org](http://www.insna.org).

## **INSNA Board Members**

President: John Skvoretz  
Vice President: Katherine Faust  
Treasurer: Laura Koehly  
Founder: Barry Wellman  
Members: Yanjie Bian, Urik Brandes, Carter Butts, Dimitris Christopoulos, Noshir Contractor, Emmanuel Lazega, David Lazer, Alessandro Lomi, and Garry Robins

## **INSNA (International Network for Social Network Analysis)**

Hardcopy circulation of Connections is sent to all members of INSNA, the International Network for Social Network Analysis, which has over 1300 members. Subscription to Connections can be obtained by registering for INSNA membership through the website: [www.insna.org](http://www.insna.org). Standard annual membership fee is US\$60 (\$40 for students). Wherever possible, items referenced in articles (such as data and software) are made available electronically through the INSNA website. In addition, the website provides access to a directory of members' email addresses, network datasets, software programs, and other electronically stored items.

## **Sunbelt Social Network Conferences**

The Sunbelt Conferences bring researchers together from all over the world to share current theoretical, methodological, and empirical findings around social networks. While previous annual conferences for INSNA members have been held in North America and Europe, conference locations will now expand to include other countries such as Australia and those in East Asia. Information on the annual Sunbelt Social Network Conferences can also be found on the INSNA website. Sunbelt XXXIV will be held in St. Pete, Florida, February 18-23, 2014.

## **Manuscript Submissions**

Submit articles to [editorconnections@gmail.com](mailto:editorconnections@gmail.com). Full manuscripts should be submitted as an MS Word document, and should not exceed 6000 words including all tables and figures. Brief Reports submissions should be under 2500 words. All images, figures and tables should be sent as separate files. Raster (photographic) images and figures should be sent in a high resolution (300ppi min.) graphics format (EPS, TIFF, or JPEG), while line work images and figures should be sent as a vector-based format (EPS or SVG). Format and style of manuscript and references should conform to the conventions specified in the latest edition of the Publication Manual of the American Psychological Association. Further instruction on submission formatting can be found on the INSNA website, <http://www.insna.org>. Manuscripts that do not follow submission criteria will be returned for revision. The journal follows a double-blind review process for research articles. Published articles are protected by both the United States Copyright Law and International Treaty provisions. All rights reserved (ISSN 0226-1776).

## **DEN (Data Exchange Network)**

We recently introduced a new section in Connections, which functions as an outlet for publishing datasets. The Data Exchange Network (DEN) has been designed with two goals: first is to build a community resource for network datasets and the second is to provide a format for citable references of datasets and instruments. Submissions should include an electronic version of the network dataset and/or instrument and a short article (not to exceed 2500 words) describing the data being submitted. All materials submitted for the DEN will be peer-reviewed to ensure the utility and usability of the data/instrument. Accepted DEN contributions will appear in the hard copy of Connections, and the data sets will be available on the INSNA website through an indexed, searchable web interface.

## INSNA Board Meeting Minutes

Friday, 24 May 2013, 12:00pm-1:30pm  
Hamburg, GERMANY

Association Manager: Julie Hewett, JulNet Solutions

Guests: Bruce Cronin, Martin Everett, David Tindall

Members Attending: George Barnett, Ulrik Brandes, Carter Butts (skyped), Dimitris Christopoulos, Mario Diani, Katherine Faust, Laura Koehly, Garry Robins, Thomas Valente, John Skvoretz, Marijtje van Duijn

Members Not Attending: David Lazar, Barry Wellman

In open session, the board heard from Bruce Cronin and Martin Everett regarding a proposal to hold Sunbelt XXXV in 2015 in Brighton UK, and from David Tindall regarding a proposal to hold it in Whistler, BC, Canada.

The Board then went into closed session and discussed the following topics:

1. Conference registration and last minute changes. It was decided to remove affiliated member designation (which currently is based on subscription to SOCNET); to keep abstract submission decoupled from conference registration; however, maintain that registration be required before the conference program is finalized. Thus, we need appropriate deadline and clarification for that. Those presenting who are not members will be required to pay a higher registration fee than members. If the primary paper presenter is unable to make it to conference, it is his/her responsibility to find someone else to present the paper and potential conflicts in the allocated presentation time.
2. The board decided that they would not endorse any research using the membership or conference registration contact information unless this is a project commissioned by the board. The database of paper presentations for INSNA is a publicly accessible database.
3. Ulrik Brandes asked that the board review a proposal for INSNA regional conferences. The board declared its willingness to sponsor INSNA regional conferences. This sets the stage for there to be a joint UK/Swiss/European network conference in years in which Sunbelt does not occur in Europe. These conferences will be synchronized such that they are not in close proximity with the annual Sunbelt meeting.
4. It was decided that there will be one main Sunbelt per year. The rotation will change, with a North American Sunbelt in alternating years. The other alternating years will be in Europe (2) and East Asia/Australia (1), for a 6 year rotation as per Mario Diani's proposal that the board accepted at its 2012 Sunbelt meeting. Were an East Asia/Australia/non European-non North American site not be available, another North American site would be pursued.
5. We reviewed proposals to host Sunbelt 2015 and beyond and make recommendations. It was decided to go to Brighton in 2015. We will solicit more proposals for a 2016 North American conference. Reservations were expressed about the cost of the Whistler BC proposal and the board encourage the proposer, David Tindall, to consider with INSNA help a proposal for Vancouver/UBC for 2016.
6. Officers (Skvoretz, Faust, Valente) led a performance review of Julnet services including conference support, web presence and other issues. Response and follow through on communications was a consistently problematic area. Some feedback was provided; however, time was short for discussion. Please send any feedback for JulNet through Skvoretz.
7. Of the candidates for board vacancies, 8 received at least 2/3rds support from current board members for a three year term (Yanjie Bian, Carter Butts, Noshir Contractor, Katherine Faust, Laura Koehly, Emmanuel Lazega, Alessandor Lomi, John Skvoretz). At the meeting of INSNA members on the 25th of May the 8 candidate Board members, along with the 4 continuing board members (Ulrik Brandes, Dimitris Christopoulos, David Lazar, Garry Robins), were presented for a confirmatory vote. Three officers were put forth and elected by acclamation – John Skvoretz (President), Katie Faust (Vice President), Laura Koehly (Treasurer).
8. A brief report on the achievements and challenges faced by the Connections editorial board was presented by the editor Dimitris Christopoulos. The requirement for resources to support an online submission platform were discussed.
9. Bylaw changes and dispensation of reserves to be discussed at the next Board meeting.

Reported by  
John Skvoretz,  
Professor of Sociology, AAAS Fellow  
President, INSNA  
Department of Sociology  
University of South Florida





## Contents

<b>“It Takes a Network”: The Rise and Fall of Social Network Analysis in U.S. Army Counterinsurgency Doctrine</b>	1
<i>David Knoke</i>	
<b>A Computationally Efficient Approximation of Beta Centrality</b>	11
<i>Zachary Neal</i>	
<b>Community Structure in Multi-Mode Networks: Applying an Eigenspectrum Approach</b>	18
<i>David Melamed, Ronald L. Breiger, and A. Joseph West</i>	
<b>Injection Drug Users’ Involvement In Drug Economy: Dynamics of Sociometric and Egocentric Social Networks</b>	24
<i>Cui Yang, Carl Latkin, Stephen Q. Muth, and Abby Rudolph</i>	
<b>Brief Reports</b>	
• <b>Social Conformity in Networks</b>	35
<i>Ian McCulloh</i>	
<b>Date Exchange Network</b>	
• Airnet 2000 Dataset	43
<i>Zachary Neal</i>	
• Multi-Relational International Trade Networks, 1965-2000	46
<i>Matthew Mahutga</i>	

## **“It Takes a Network”: The Rise and Fall of Social Network Analysis in U.S. Army Counterinsurgency Doctrine**

---

**David Knoke**

*Department of Sociology, University of Minnesota  
Minneapolis, Minnesota*

### **Abstract**

During the Iraq and Afghanistan Wars, a group of warrior-thinkers developed a new U.S. Army counterinsurgency (COIN) doctrine to fight modern “jihadist” insurgencies. Drawing heavily on social network analysis ideas, COIN principles emphasized population protection and organizational learning and adaptation. As implemented in Iraq by General David Petraeus, the doctrine greatly reduced intercommunal violence although other factors also contributed. But, COIN in Afghanistan under General Stanley McChrystal was unsuccessful in ending the Taliban insurgency. Although the Obama Administration substantially diminished the U.S. Army’s counterinsurgency capabilities, social network analytic ideas persist in military policy and practices.

### **Authors**

*David Knoke* is professor of sociology at the University of Minnesota in Minneapolis, MN. He teaches about and conducts research on diverse social networks, including intraorganizational, interorganizational, health care, economic, financial, terrorist, and counterterror networks. His most recent book is *Economic Networks* (2012 Polity Press).

### **Notes**

This article began as a keynote address to the 5th Annual Political Networks Conference and Workshops, July 13-16, 2012, at the University of Colorado in Boulder, CO. The author thanks an anonymous reviewer for some valuable suggestions.

*Please email all correspondence to: [knoke001@umn.edu](mailto:knoke001@umn.edu).*



## 1. Introduction

*"It takes a network to defeat a network."*  
Gen. Stanley A. McChrystal (2011)

United States Navy Seal Team Six killed Osama bin Laden after social network analysis methods discovered his secret location. The Central Intelligence Agency had conducted a decade-long manhunt to find and capture or kill the Al-Qaida mastermind behind the September 11 attacks. Because bin Laden never used the Internet or telephone, all his communications with top Al-Qaida commanders relied on a courier who relayed messages to and from bin Laden's hideout in Abbottabad, Pakistan. The breakthrough came with the discovery of the courier's true identity – a Pakistani whose nom-de-guerre was Abu Ahmed al-Kuwaiti (Goldman & Apuzzo, 2011). In 2010, a CIA wiretap overheard another Al-Qaida operative conversing with al-Kuwaiti. (Disputed is the extent to which the torture of detainees also yielded relevant information, as implied in the action thriller *Zero Dark Thirty* [Rodriguez, 2013].) CIA agents tracked Kuwaiti's white S.U.V., its spare-wheel cover painted with a white rhino, to the Abbottabad compound. After months of aerial surveillance, the CIA concluded that Kuwaiti and his brother lived there, and speculated that a third man – "The Pacer," who never went outside the compound walls – was bin Laden. On the night of May 2, 2011, SEAL Team Six raided the compound and killed bin Laden, the courier, and his brother. (For details about the raid, see Schmidle, 2011; Bergen, 2012; Owen, 2012).

The hunt for bin Laden epitomized how intelligence organizations applied network analytic methods to identify and map connections among members of terrorist and insurgent organizations, and how counterinsurgency forces working as a network used that information in field operations to disrupt and destroy them. In the early twenty-first century, network ideas infused new U.S. Army counterinsurgency doctrine: "In bitter, bloody fights in both Afghanistan and Iraq, it became clear to me and to many others that to defeat a networked enemy we had to become a network ourselves" (McChrystal, 2011). This article analyzes the rise and fall of this new counterinsurgency doctrine and the persistence of social network analysis in military policy and practices.

## 2. Classical and Modern Insurgencies

Analysts of asymmetric warfare identified major differences between classical "Maoist" insurgencies and modern "jihadist" insurgencies (Muckian, 2006). Mao Zedong's revolutionary strategy – expressed in his aphorism "the guerilla must move among the people as a fish swims in the sea" – emphasized a hierarchical military command structure fully integrated with a parallel political hierarchy. This military-political structure mobilizes and indoctrinates the local populace in the goals of the revolution. Tactically, guerilla armies operate in small-unit formations, depend on support from local populations, and try to bleed the enemy forces through ambushes and raids on police stations and military barracks. Another Maoist saying is apt: "The enemy advances, we retreat; the enemy camps, we harass; the enemy tires, we attack; the enemy retreats, we pur-

sue" (Zedong, 1965, p. 124). Insurgents fight to seize and hold territory, eliminate local agents of the national government, and replace them with revolutionary cadres. Ultimately, the insurgency builds sufficient popular support and military strength to defeat the old regime by conventional warfare on the battlefield. Many liberation struggles applied the Maoist strategy from the 1940s through the 1960s: successfully in China, Cuba, Vietnam, and Cambodia, and unsuccessfully in Che Guevara's 1965 Congo and 1966 Bolivian campaigns, and the Nepali People's War at the end of the twentieth century.

Classical insurgencies begat classical counterinsurgency doctrines that aimed to decapitate an insurgency's politico-military leaders and to dry up the sea of the people (a.k.a., winning-hearts-and-minds). During the 1948-1960 Malayan Emergency led by the Malayan Communist Party, the British forcibly resettled a half-million peasants of predominantly Chinese ancestry into 450 guarded "New Villages." This strategy denied the guerillas access to information and resources from a sympathetic population (Thompson, 1966). It's widely regarded as one of a few unambiguously successful counterinsurgency campaigns by a foreign occupying power (Nagl, 2005). Far less successful were two programs implemented by the United States during the 1955-1975 Vietnam War. The Strategic Hamlet Program, partly modeled on the British experience in Malaya, forcibly resettled more than eight million peasants into 7,000 villages. But, the South Vietnamese central government provided insufficient security against attacks by Viet Cong insurgents and the program collapsed after the 1963 South Vietnamese military coup against the Diem regime. The CIA's Phoenix Program in the late-1960s sought to "neutralize" – capture, convert, or assassinate – suspected Viet Cong cadres and their civilian sympathizers (Andrade, 1990). Carried out by local South Vietnamese militias and police, it tortured and killed tens of thousands of suspects before a U.S. congressional backlash against its abuses shut down the Phoenix Program (Valentine, 1990).

Modern jihadist insurgencies and terrorist campaigns differ from their Maoist predecessors in basic organizational structures and strategies. In place of vertically integrated hierarchies, jihadis assemble in continually shifting networks of militants. Groups consist of numerous small, self-organized cells, as few as two or three people, and the complete network exhibits low density and low social cohesion (Knoke, 2012). A decentralized network has no core leadership and no political cadre that exerts control over the cells' militant actions. Jihadis make effective use of modern information technologies – cell phones, Internet, Websites, video clips – to recruit participants, propagandize their successful attacks, mobilize popular support, plan tactical operations, and coordinate attacks on occupying forces. The insurgency has no explicit political program for mobilizing the populace, and makes no attempt to capture and hold a territory. The insurgency's immediate aim is to drive out the foreign occupying forces by inflicting such high levels of injury and death that democratic governments will be forced to withdraw. But, the insurgents have no clearly articulated long-term goals for their nation.

At the start of the Iraq Occupation in 2003, the U.S.-led Coalition Forces deployed a counterinsurgency strategy pre-

mised on a belief that they faced a conventional Maoist insurgency. They hunkered inside large, heavily fortified forward operating bases, isolated from population centers, from where patrols emerged daily and to which they retired by nightfall. Thus, insurgent cells had the nocturnal run of urban neighborhoods and rural areas, to collect information, extort resources, and plant improvised explosive devices (IEDs) that wreaked bodily havoc on the next morning's foot patrols. The counterinsurgency attempted to capture and kill the leaders of a dozen or more major militant organizations, including Al-Qaida in Iraq, Ansar al-Islam, Badr Brigade, and Mahdi Army. But, decentralized structures rendered a decapitation strategy fruitless. Self-organizing networks were highly resilient to loss of militant foot-soldiers, because replacements were easily recruited. The tenuous connections among cells meant that picking up and interrogating individuals could not unravel the complete network.

### 3. Minting the New COIN Doctrine

On May 1, 2003, President Bush landed on the carrier USS Abraham Lincoln and saluted the troops at the end of combat operations under a banner proclaiming "Mission Accomplished." Soon after, the Iraq Occupation began to spiral into a morass of death squads, ethnic cleansing, banditry, terrorism, and religious insurgency that steadily worsened year after year. By February, 2007, the Iraq National Intelligence Estimate foresaw a potential civil war:

*The Intelligence Community judges that the term "civil war" does not adequately capture the complexity of the conflict in Iraq, which includes extensive Shia-on-Shia violence, al-Qa'ida and Sunni insurgent attacks on Coalition forces, and widespread criminally motivated violence. Nonetheless, the term "civil war" accurately describes key elements of the Iraqi conflict, including the hardening of ethno-sectarian identities, a sea change in the character of the violence, ethno-sectarian mobilization, and population displacements.*

(ThinkProgress, 2007)

Such projections of looming defeat, redolent of the 1960s U.S. debacle in Vietnam, had already impelled some members of the military to formulate a new doctrine for fighting modern jihadist insurgencies, one which made extensive use of social network analytic ideas.

In late 2005, U.S. Army Lieutenant General David Petraeus, who had commanded the storied 101st Airborne Division during the 2003 conquest of Baghdad, was appointed commander of Fort Leavenworth, Kansas, and its Army Combined Arms Center (CAC). Among other duties, he oversaw the preparation and publication of U.S. Army/Marine Counterinsurgency Field Manual 3-24, the first publication in two decades devoted to the topic. FM3-24 was the product of a team of "warrior-thinkers" – assembled by Petraeus and Marine Lieutenant General James "Mad Dog" Mattis – that culled lessons from historical "small wars" and past counterinsurgency successes and failures. They invited 80 experts to a February

2006 conference at Fort Leavenworth, including military and civilian participants "ranging from veterans of Vietnam and El Salvador to human rights advocates, who deconstructed the draft chapters and made the final product stronger" (Nagl, 2010, p. 118). The revised final draft of FM3-24, coauthored by Petraeus and Marine Lieutenant General James F. Amos, was published in December 2006, and became an instant best-seller. While commander of Fort Leavenworth, Petraeus integrated FM3-24 lessons into classroom teaching and training exercises for officers at CAC military schools.

The new COIN doctrine was "built around two big ideas: first, that protecting the population was the key to success in any counterinsurgency, and second, that to succeed in counterinsurgency, an army has to be able to learn and adapt more rapidly than its enemy" (Nagel, 2010 p. 118). These twin pillars are population-centric and enemy-centric COIN, respectively. The first pillar emphasized a clear-hold-build approach to population protection. Because "the cornerstone of any COIN effort is establishing security for the civilian populace," commanders must move quickly to shift from combat operations to building law-enforcement institutions such as police, courts, and penal facilities in the host nation (Petraeus & Amos, 2006, p. 1-23-24). Other responsibilities include provision of essential services such as water and medical care, and "sustainment of key social and cultural institutions" (p. 2-2). In effect, population-centric COIN depends heavily on nation-building activities that gain legitimacy and local support for the national government. Many activities implicitly require the military to form network connections with the populace and interorganizational ties to host-nation institutions. COIN leadership "can design an operation that promotes effective collaboration and coordination among all agencies and the affected populace" (p. 2-4).

Nine Zen-like statements summarizing COIN principles and imperatives – labeled "paradoxes of counterinsurgency" – described how the new doctrine differed from "the traditional American view of war" (p. 1-26). The first paradox, which emphasized networking to achieve the ultimate success of protecting the populace, was: "Sometimes, the more you protect your force, the less secure you may be."

*If military forces remain in their compounds, they lose touch with the people, appear to be running scared, and cede the initiative to the insurgents. Aggressive saturation patrolling, ambushes, and listening post operations must be conducted, risk shared with the populace, and contact maintained. ... These practices ensure access to the intelligence needed to drive operations. Following them reinforces the connections with the populace that help establish real legitimacy. (p. 1-27)*

Another paradox, "Some of the best weapons for counterinsurgents do not shoot," argued that political, social, and economic programs are often more important than conventional firepower for undermining an insurgency. "Arguably, the decisive battle is for the people's minds ... While security is essential to setting the stage for overall progress, lasting victory comes from a vibrant economy, political participation, and

restored hope” (p. 1-27). For more on COIN paradoxes, see Cohen et al. (2006).

In the enemy-centric pillar, organizational learning required that soldiers and marines study FM3-24 and its source materials before deployment, then “apply what they have learned through study and experience, assess the results of their actions, and continue to learn during operations” (p. x). Among the most important skills to learn is social networking, both as a means to build rapport with the populace and as a method to detect and destroy insurgent organizations:

*This requires living in the AO [area of operations] close to the populace. Raiding from remote, secure bases does not work. Movement on foot, sleeping in villages, and night patrolling all seem more dangerous than they are – and they are what ground forces are trained to do. Being on the ground establishes links with the local people. They begin to see Soldiers and Marines as real people they can trust and do business with, rather than as aliens who descended from armored boxes. (Appendix A-4)*

FM3-24 identified enemy networks as:

*A tool available to territorially rooted insurgencies, such as the FARC in Colombia. Other groups have little physical presence in their target countries and exist almost entirely as networks. Networked organizations are difficult to destroy. In addition, they tend to heal, adapt, and learn rapidly. However, such organizations have a limited ability to attain strategic success because they cannot easily muster and focus power. The best outcome they can expect is to create a security vacuum leading to a collapse of the targeted regime’s will and then to gain in the competition for the spoils. However, their enhanced abilities to sow disorder and survive present particularly difficult problems for counterinsurgents. (p. 1-17)*

FM3-24 Appendix B, “Social Network Analysis and Other Analytical Tools,” described “social network analysis, a powerful threat evaluation tool” to help commanders and staff “understand the operational environment” (p. B-1). It explained basic network concepts and measures for identifying and portraying details of insurgent network structures. Figure 1 (as it appears in FM-24, Fig. B-7) illustrated how a well-executed COIN, by decreasing an insurgent network’s density, could erode its ability to conduct coordinated attacks, which “means the group is reduced to fragmented or individual-level attacks” (p. B-12). High-density networks “require only the capture of one highly connected insurgent to lead counterinsurgents to the rest of the group. So while high-network-density groups are the most dangerous, they are also the easiest to defeat and disrupt.” The Appendix went into great detail about network concepts such as core-periphery, centrality, diameter, and hubs, and how they could be used to identify key actors.

A sidebar story described how Saddam Hussein was captured in December 2003 after months of painstaking intelligence gathering. Analysts constructed link diagrams showing people related to Hussein by blood or tribe:

*Each day another piece of the puzzle fell into place. Each led to coalition forces identifying and locating more of the key players in the insurgent network ... Analysts traced trends and patterns, examined enemy tactics, and related enemy tendencies to the names and groups on the tracking charts. This process involved making continual adjustments to the network template and constantly determining which critical data points were missing. (Petraeus & Amos 2006, p. B-14)*

Using up-to-date network diagrams, “commanders then designed a series of raids to capture key individuals and leaders from the former regime who could lead counterinsurgents to him” (p. B-14). The cycle continued, “eventually leading coalition forces into Hussein’s most trusted inner circle and finally

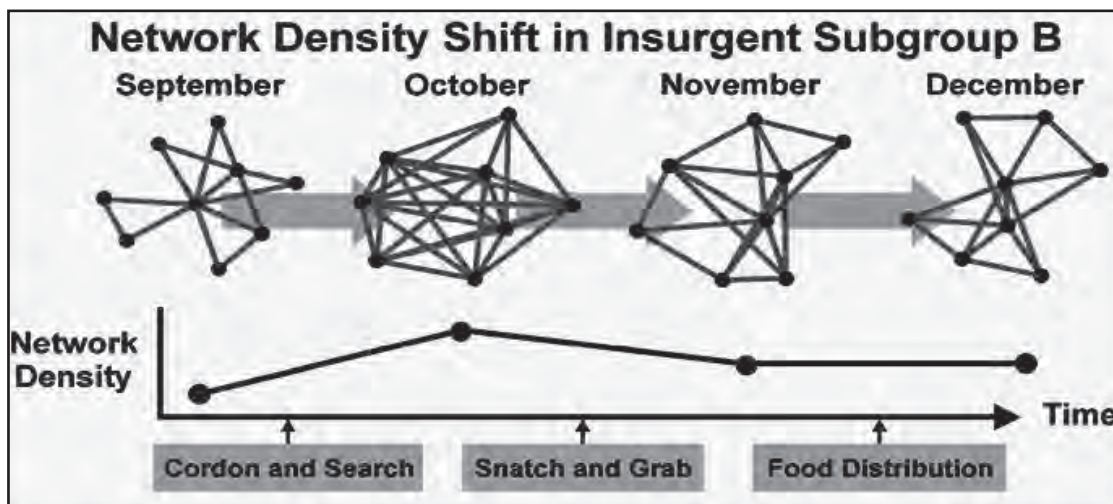


Figure 1. Example of changes to tactics on density shift.

to Hussein’s capture.” For detailed explications of these procedures, see publications by Col. Brian Reed, a member of the Saddam Hussein task force and contributor to FM3-24 Appendix B (Reed, 2006; Reed & Segal, 2006; Reed, 2007).

Other COIN intelligence tools related to social network analysis included activities matrices (two-mode data) and association matrices (one-mode data). In conclusion, “SNA can help commanders determine what kind of social network an insurgent organization is. That knowledge helps commanders understand what the network looks like, how it is connected, and how best to defeat it” (p. B-17).

**4. COIN in Iraq**

Within one year after publishing FM3-24, Petraeus got his chance to put the new COIN doctrine to a field test. Conditions in Iraq had deteriorated so much that the Republican Party lost its majorities in both the Senate and House of Representatives in the November 6, 2006, mid-term elections. Acknowledging “it was a thumping,” two days later President Bush fired Secretary of Defense Donald Rumsfeld and nominated a former CIA Director, Robert Gates, to replace him. The Bush Administration conducted a review of strategic options, and in January 2007, President Bush announced “the surge” – a deployment of more than 20,000 additional soldiers and marines to Iraq with a new mission: “to help Iraqis clear and secure neighborhoods, to help them protect the local population, and to help ensure that the Iraqi forces left behind are capable of providing the security that Baghdad needs” (Bush, 2007). His population-protection rhetoric was straight from the new COIN doctrine, and among other personnel changes, Bush appointed newly promoted four-star General Petraeus to command the Coalition Forces. Petraeus issued counterinsurgency guidance to the Multi-National Force-Iraq which reiterated many of the COIN principles in FM3-24. A widely disseminated diagram, “Anaconda Strategy vs. AQI” (Al-Qaida in Iraq), drew an implicit parallel to a

U.S. Civil War plan for the Union to suffocate the Confederacy by blockading Southern ports and cutting the South in two by advancing down the Mississippi River (see Figure 2).

Surge troops were dispersed among 30 neighborhoods in Baghdad, to live in “joint security stations” which they shared with Iraqi military and police forces. Their long-term presence in local communities was designed to build trust relations with the residents, and gain their cooperation for discovering arms caches, identifying IED makers, and rooting out insurgents. “At the same time, American forces launched an all-out assault on Al Qaeda strongholds that ringed the capital” (Filkins, 2012). Although violence against civilians and military peaked in the early months of the surge, by September 2007, Petraeus testified to Congress that “the military objectives of the surge are, in large measure, being met” and that COIN operations had greatly reduced sectarian violence between Sunni and Shia (Cloud & Shanker, 2007). Violence declined dramatically through the following year, as the surge came to a conclusion in July 2008. Although President Obama withdrew the last U.S. troops in December 2011, a low-grade Iraqi insurgency persists mainly in Sunni insurgent attacks on civilians and the Shia-dominated national government.

Three events, occurring before or during the 2007-08 surge, may have contributed to the apparent successful implementation of the COIN doctrine. First, preceding years of ethnic-cleansing in many Baghdad neighborhoods greatly diminished opportunities for further intercommunal violence. Second, the U.S. military began to pay Sunni tribal sheiks in Anbar Province to stop cooperating with Al-Qaida against the Coalition and to hire a hundred thousand former insurgents as local security forces.

*Al-Qaeda in Iraq had made a strategic mistake in the province, overplaying its hand. Its members had performed forced marriages with women from local tribes, taken over hospitals, used mosques for*



Figure 2. Anaconda Strategy vs. AQI.

*beheading operations, mortared playgrounds and executed citizens, leaving headless bodies with signs that read, “Don’t remove this body or the same thing will happen to you.” The sheer brutality eroded much of the local support for al-Qaeda in Iraq. (Woodward, 2008)*

Some analysts viewed this “Sunni Awakening” movement as more decisive than the troop surge in bringing relative stability to Iraq (Litchfield, 2010; Coulter, 2010; Jones, 2012 p. 239-260). A third contributing factor was a series of top-secret operations in Spring, 2007, conducted jointly by “fusion cells” of U.S. special forces and intelligence agents, “to locate, target and kill key individuals in groups such as al-Qaeda in Iraq, the Sunni insurgency and renegade Shia militias, or so-called special groups” (Woodward, 2008). A major proponent of this “collaborative warfare” was Lt. Gen. Stanley McChrystal, who soon applied social network methods to the deteriorating War in Afghanistan.

## 5. COIN in Afghanistan

While U.S. military efforts focused on Iraq, the conflict in Afghanistan steadily intensified following initial success in driving the Taliban from power and installing a national government in Kabul led by President Hamid Karzai. Resurgent Taliban forces intimidated villagers to provide support and resources, and deployed roadside IEDs to attack convoys of the International Security Assistance Force (ISAF), the occupying U.S. and NATO forces. Could the lessons learned about COIN in Iraq be replicated in Afghanistan to turn that war around? Petraeus recognized “enormous differences” between the two war theaters:

*You have to apply it [counterinsurgency] in a way that is culturally appropriate for Afghanistan. For example, a key strategy shift that accompanied the troop surge in Iraq – in which U.S. troops lived within the Iraqi communities they helped to secure – won’t necessarily work in Afghanistan. You don’t move into a village in Afghanistan the way that we were able to move into neighborhoods in Iraq. You have to move on the edge of it, or just near it, but you still have to have a persistent security presence. (Miles, 2009)*

In June 2009, General Stanley McChrystal took command of ISAF. He had previously commanded the Joint Special Operations Command in Iraq, which captured Saddam Hussein and killed Abu Musab al-Zarqawi, leader of Al-Qaida in Iraq. Through those experiences, McChrystal became a convert to the new Army COIN doctrine. He depicted the Taliban as “more network than army, more a community of interest than a corporate structure” (2011). Hence, successful COIN in Afghanistan necessitated creating an opposing network – connecting intelligence analysts, drone operators, and combat teams – capable of rapidly sharing real-time information gathered during night raids on insurgents:

*In bitter, bloody fights in both Afghanistan and Iraq, it became clear to me and to many others that to defeat a networked enemy we had to become a network ourselves. We had to figure out a way to retain our traditional capabilities of professionalism, technology, and, when needed, overwhelming force, while achieving levels of knowledge, speed, precision, and unity of effort that only a network could provide. We needed to orchestrate a nuanced, population-centric campaign that comprised the ability to almost instantaneously swing a devastating hammer blow against an infiltrating insurgent force or wield a deft scalpel to capture or kill an enemy leader. (McChrystal, 2011)*

McChrystal’s application of COIN also stressed protecting the Afghan populace, implementing reconstruction and development projects, and strengthening the local Afghan government’s legitimacy so villagers would withhold support from the insurgency.

In September 2009, McChrystal’s report to the Pentagon assessing the war’s bleak prospects was leaked to the press, a blatant attempt to influence policy that some critics charged was insubordinate (Woodward, 2009). McChrystal argued that, without an additional 40,000 troops and the application of a genuine counterinsurgency strategy, the mission “will likely result in failure.” Following a White House strategy review dominated by the military and its supporters, President Obama in December 2009 announced a surge of 30,000 U.S. troops into Afghanistan, and set 18 months as the deadline for their withdrawal. Under those resource constraints, McChrystal’s effort to implement COIN ultimately failed. Some U.S. field commanders simply ignored orders to protect civilians, giving top priority instead to conventional search-and-destroy operations against the Taliban (Chandrasekaran, 2012, p. 147-169). Afghanistan simply never experienced a surge of skilled social network analysts compared to the Iraq War.

A major obstacle to successful COIN was the incompetence and corruption of the Karzai regime, which was either unwilling or unable to extend its writ much beyond Kabul (Filkins, 2012; McChrystal, 2013; Kaplan, 2013). ISAF alliances with provincial militias, which supplied them with weapons and money, “led to counter-productive results such as the strengthening of local Power Brokers and the weakening of the government in Kabul” (Gauster, 2008, p. 11). The plight of the Provincial Reconstruction Teams (PRTs) in Afghanistan revealed another hindrance. From the early years of the Afghan War, the ISAF deployed two dozen PRTs as part of its population-protection strategy. PRTs consisted of military officers, State Department diplomats, and technical experts from USAID and the Department of Agriculture. A military officer commanded each PRT staff of 80 to 250, only a handful of whom were civilians. PRT goals were to improve local security; directly fund and assist reconstruction projects, such as schools and clinics; and extend the legitimacy of the Afghan central government into the provinces. But, differences in civilian and military cultures, unclear lines of authority, and clashes among agencies hindered the integration and effectiveness of

the PRT networks (Luehrs, 2009; Fritsch, 2012). At the Munich Security Conference in February 2011, President Karzai criticized the PRTs, along with private security firms, as impediments to extending the central government's authority into the countryside. One analyst concluded that, as ISAF withdraws by 2014, a gradual transition of PRT functions and international funding to the Afghan government may be "actually better for the long-term health of Afghanistan, even if it contributes to aggregate corruption" (Foust, 2011). In the absence of a credible and dependable host-nation partner, even the most incisively executed military counterinsurgency stood little chance.

In June 2010, just one year after taking command, McChrystal resigned when *Rolling Stone* reported disrespectful remarks his aides made about the Obama Administration. President Obama replaced him with Petraeus, who, despite recycling his counterinsurgency guidelines from Iraq, now emphasized a relentless kill-or-capture campaign against the Taliban and Al-Qaida remnants:

*Pursue the enemy relentlessly. Together with our Afghan partners, get our teeth into the insurgents and don't let go. When the extremists fight, make them pay. Seek out and eliminate those who threaten the population. Don't let them intimidate the innocent. Target the whole network, not just individuals.* (Petraeus, 2010)

Petraeus ramped up night raids, with black helicopters dropping Special Ops forces into villages to capture or kill Taliban commanders and financiers (Chandrasekaran, 2012 p. 148-169). Increased Predator drone strikes inside Afghanistan caused heavy civilian casualties ("collateral damage"), incensing President Karzai. Ironically, the "counterterrorism policy of raids and air strikes that Petraeus and other commanders had derided in the 2009 White House strategy review had become the military's principal tool to weaken the insurgency" (p. 278). Although Petraeus denied abandoning his signature approach, evidence indicated that most intelligence about insurgents came from signal intercepts rather than from tips by the local populace. After a year as ISAF commander, Petraeus departed to head the CIA, supervising its drone war on both sides of the Af-Pak border. The U.S. troop surge wound down, yet the Taliban still lurk in the countryside, biding their time until the occupiers leave in 2014, if not sooner.

## 6. Persistence of SNA in Military Policy and Practices

COIN doctrine and its applied network methods came full-circle in just half a decade. Itself an insurgency by dissident military intellectuals against conventional military thinking, COIN sought radically to remake the American way of war (Kaplan, 2013). Its advocate generals, Petraeus and McChrystal, briefly seized the strategic high ground in urban Iraq, backed by an American president desperately trying to avoid another military debacle, only to crash and burn under the harsh realities of insurgency in mountainous Afghanistan. The brief COIN renaissance they wrought finally evaporated in the heat of the 2012 presidential election debates when President Obama de-

clared, "after a decade of war, it's time to do some nation building here at home" (Jackson, 2012). He proposed a Pentagon budget that cuts \$487 billion over 10 years, by downsizing the Army and Marine Corps to smaller forces capable of only "limited counterinsurgency." Instead, military efforts will shift to "intelligence, surveillance, reconnaissance, counterterrorism, countering weapons of mass destruction, and the ability to operate in environments where adversaries try to deny us access" (Ackerman, 2012). Drones and commandos will hunt, capture, and kill terrorists in Pakistan, Yemen, Somalia, and other un-governed territories.

Military strategists, policymakers, and historians heatedly debate the relevance of COIN doctrine for the Iraq and Afghanistan Wars (Ucko, 2008; Paul & Clarke, 2011; Branch & Wood, 2012; Gilmore, 2012; Kurtulus, 2012; Jones et al., 2012). For example, Colonel Gian Gentile criticized the closed bureaucratic process by which population-centric COIN "came to dominate how the Army thinks about war without a serious professional and public debate over its efficacy, practicality, and utility" (Gentile, 2010, p. 116). He charged that the military "bought into a doctrine for countering insurgencies that did not work in the past, as proven by history, and whose efficacy and utility remain highly problematic today." Gentile called for *FM3-24* to be "deconstructed" through an open forum process that would provide a "better and more complete operational doctrine for counterinsurgency" (p. 117). In rebuttal, John Nagl (2010), a contributor to *FM3-24*, acknowledged its heavy reliance on "classical" counterinsurgency theory, but argued that "the differences between previous and current insurgencies are overstated." He asserted that the Army's subsequent development of the fifteenth edition of its capstone doctrine, *Operations Field Manual 3-0* (Dempsey, 2008), was "a revolutionary departure from past doctrine", produced through a more rigorous internal review than conducted for *FM3-24* (Nagl, 2010:119). Other recent manuals were also published after open development processes, "making the preparation of doctrine less about traditional practices and more about constant learning and adaptation based on current experience and collaboration with a broad group of concerned partners" (p. 120).

Social network analytic ideas pervade *FM3-0*, which emphasizes networking among military forces to defeat enemy networks. The Army faces "hybrid threats" from "diverse and dynamic combination of regular forces, irregular forces, and criminal elements, or a combination of these forces and elements" (p. 1-5). To fight such foes, the Army needs "flexible mission command networks and systems" (p. 3-21) to "enable the art of command and science of control" (p. 4-6). "Networks are key enablers to successful operations," and network systems provide "synthesized information so leaders can make informed decisions without being overburdened" (p. 6-6). Control of cyberspace is essential both for effective command of Army forces and as "a venue to attack enemy networks and systems" (p. 6-15). "Cyber warfare uses cyber exploitation, cyber attack, and cyber defense in a mutually supporting and supported relationship with network operations and cyber support" (p. 6-21). The implementation of these evolving network-centric principles will undoubtedly be battlefield-tested in future

combat operations of the U.S. military around the world and in cyberspace.

Social network analysis has traction at the Joint Improvised Explosive Device Defeat Organization (JIEDDO), founded in 2006 by the Pentagon to reduce the strategic influence of IEDs. Its proclaimed mission is to “defeat the device, attack the network, and train the force.” At JIEDDO training facilities, “units are shown how to identify, map and target insurgent and terrorist networks, while positively influencing friendly and neutral populations and networks” (JIEDDO, 2013). However, a recent critical assessment concluded that, although the organization excelled in training and defeating IEDs, it still lagged “in providing necessary information to facilitate attack-the-network operations” (Morganthaler & Giles-Summers, 2011, p. v). To compensate for this deficiency, a “focused SNA approach would provide the framework to build and sustain knowledge of dark networks across time and unit rotations” (p. 40). Ultimately, strategic success depends on a better understanding of the broader social contexts which generate insurgency and terrorism.

A major shortcoming of the new COIN doctrine was its failure to institutionalize social network analysis training and education following initial success in Iraq. No military school exists for SNA and these methods are not taught at military intelligence schools. The Army Additional Skill Identifier lists no occupational code for a social network analyst, which would explicitly assign and train personnel in those skills (US Military About, 2013). (In contrast, social network analyst jobs are regularly advertised by civilian security and intelligence firms, such as RAND Corporation, NEK, SAIC, Harrison Corporation, Chenega, Acclaim Technical Services.) As a consequence, too few people in the military appreciate the potential value of network analytic ideas and understand how best to use them in counterinsurgency operations. The small number of staff who know network analysis methods, and who can use them to achieve tactical successes in the field, soon rotate out of theater before senior officers comprehend the bases of those successes. The failure to institutionalize the new COIN and its SNA methods is part of a larger misapprehension by the military and intelligence community about the complexity of societal conditions and social problems in developing countries where future conflicts loom. But that’s another story for another time.

## References

- Ackerman, Spencer (2012, January 5). Obama’s New Defense Plan: Drones, Spec Ops and Cyber War. *Wired*. Retrieved January 29, 2013, from <http://www.wired.com/dangerroom/2012/01/pentagon-asia-strategy>
- Andrade, Dale (1990). *Ashes to Ashes: The Phoenix Program and the Vietnam War*. Lexington, MA: Heath.
- Bergen, Peter L. (2012). *Manhunt: The Ten-Year Search for binLaden from 9/11 to Abbottabad*. New York, NY: Crown Publishing Group.
- Branch, Daniel, & Wood, Elisabeth Jean (2010). Revisiting Counterinsurgency. *Politics & Society*, 38, 3-14.
- Bush, George W. (2007, January 10). “President’s Address to the Nation.” Washington Office of the Press Secretary. Retrieved January 5, 2013, from <http://georgewbush-whitehouse.archives.gov/news/releases/2007/01/20070110-7.html>
- Chandrasekaran, Rajiv. (2012). *Little America: The War within the War for Afghanistan*. New York, NY: Knopf.
- Cloud, David S. & Shanker, Thom (2007, September 11). Petraeus Warns Against Quick Pullback in Iraq. *New York Times*. Retrieved January 5, 2013, from [http://www.nytimes.com/2007/09/11/washington/11policy.html?pagewanted=all&\\_r=0](http://www.nytimes.com/2007/09/11/washington/11policy.html?pagewanted=all&_r=0)
- Cohen, Eliot, Crane, Conrad, Horvath, Jan, & Nagl, John (2006, March-April). Principles, Imperatives, and Paradoxes of Counterinsurgency. *Military Review*, 49-53. Retrieved January 2, 2013, from <http://www.au.af.mil/au/awc/awcgate/milreview/cohen.pdf>
- Coulter, Sean D. (2010). Sandals on the Ground: Why the Awakening, Not the Surge, was Decisive in Iraq. Newport, RI: Naval War College, Joint Military Operations Department. Retrieved January 5, 2013, from <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA525326>
- Dempsey, Marti E. (2008). *Operations Field Manual 3-0*. Washington D.C.: United States Army. Retrieved January 10, 2013, from <http://www.fas.org/irp/doddir/army/fm3-0.pdf>
- Filkins, Dexter (2012, December 17). General Principles: How Good Was David Petraeus? *New Yorker*. Retrieved December 25, 2012, from [http://www.newyorker.com/arts/critics/atlarge/2012/12/17/121217crat\\_atlarge\\_filkins](http://www.newyorker.com/arts/critics/atlarge/2012/12/17/121217crat_atlarge_filkins)
- Foust, Joshua (2011). Actually, Karzai is Right about PRTs. *Foreign Policy*. Retrieved January 28, 2013, from [http://afpak.foreignpolicy.com/posts/2011/02/08/actually\\_karzai\\_is\\_right\\_about\\_prts](http://afpak.foreignpolicy.com/posts/2011/02/08/actually_karzai_is_right_about_prts)
- Fritsch, Jocelyn (2012). *Understanding U.S. Civil-Military Cooperation in the U.S. Provincial Reconstruction Teams (PRTs) in Afghanistan* (Doctoral dissertation). Minneapolis, MN: University of Minnesota. Retrieved January 28, 2013, from [http://conservancy.umn.edu/bitstream/142308/1/Fritsch\\_umn\\_0130E\\_13216.pdf](http://conservancy.umn.edu/bitstream/142308/1/Fritsch_umn_0130E_13216.pdf)
- Gauster, Markus (2008). Provincial Reconstruction Teams in Afghanistan: An Innovative Instrument of International Crisis Management Being Put to the Test. *Occasional Paper Series No. 16*. Garmisch-Partenkirchen, Germany: George C. Marshall Center. Retrieved January 28, 2013, from [http://www.humansecuritygateway.com/documents/ECSS\\_Afghanistan\\_provincialreconstructionteams.pdf](http://www.humansecuritygateway.com/documents/ECSS_Afghanistan_provincialreconstructionteams.pdf)
- Gentile, Gian (2010). Time for the Deconstruction of Field Manual 3-24. *Joint Forces Quarterly*, 58, 116-117. Retrieved January 2, 2013, from [http://www.ndu.edu/press/lib/images/jfq-58/JFQ58\\_116-117\\_Deconstruction.pdf](http://www.ndu.edu/press/lib/images/jfq-58/JFQ58_116-117_Deconstruction.pdf)
- Gilmore, Jonathan (2012). A Kinder, Gentler Counter-Terrorism: Counterinsurgency, Human Security and the War on Terror.” *Security Dialogue*, 42, 21-37.

- Goldman, Adam, & Apuzzo, Matt (2011, May 3). Phone Call by Kuwaiti Courier Led to bin Laden. *Associated Press, San Diego Union Tribune*. Retrieved January 1, 2013, from [www.uniontribune.org/news/2011/may/02/phone-call-by-kuwaiti-courier-led-to-bin-laden](http://www.uniontribune.org/news/2011/may/02/phone-call-by-kuwaiti-courier-led-to-bin-laden)
- Jones, David Martin, Smith, M.L.R., & Stone, John (2012). Counter-COIN: Counterinsurgency and the Preemption of Strategy. *Studies in Conflict & Terrorism*, 35, 597-617.
- Jackson, David (2012, October 22). Live Blog: Obama vs. Romney on Foreign Policy. *USA Today*. Retrieved January 17, 2013, from <http://www.usatoday.com/story/onpolitics/2012/10/22/obama-romney-live-blog-foreign-affairs-debate-boca-raton/1650581/>
- Joint Improvised Explosive Device Defeat Organization (JIEDDO). (2013). *Train the Force*. Retrieved February 3, 2013, from <https://www.jieddo.mil/train.aspx>
- Jones, Seth G. (2012). *Hunting in the Shadows: The Pursuit of Al Qaeda since 9/11*. New York, NY: Norton.
- Kaplan, Fred (2013). *The Insurgents: David Petraeus and the Plot to Change the American Way of War*. New York, NY: Simon and Schuster.
- Knoke, David (2012). Social Networks and Terrorism. In Maytham Safar & Khaled A. Mahdi (Eds.), *Social Networking and Community Behavior Modeling: Qualitative and Quantitative Measures* (pp. 232-246). Hershey, PA: IGI Global.
- Kurtulus, Ersun N. (2012). The New Counterterrorism: Contemporary Counterterrorism Trends in the United States and Israel. *Studies in Conflict & Terrorism*, 35, 37-58.
- Litchfield, John D. (2010). *Unconventional Counterinsurgency: Leveraging Traditional Social Networks and Irregular Forces in Remote and Ungoverned Areas*. Fort Leavenworth, KS: School of Advanced Military Studies.
- Luehrs, Christoff (2009). Provincial Reconstruction Teams: A Literature Review. *Prism*, 1, 95-102.
- McChrystal, Stanley A. (2011, March-April). It Takes a Network: The New Front Line of Modern Warfare. *Foreign Policy*. Retrieved April 24, 2012, from [http://www.foreignpolicy.com/articles/2011/02/22/it\\_takes\\_a\\_network?page=full](http://www.foreignpolicy.com/articles/2011/02/22/it_takes_a_network?page=full)
- McChrystal, Stanley A. (2013). *My Share of the Task: A Memoir*. New York, NY: Penguin Portfolio.
- Miles, Donna (2009, April 22). Petraeus Discuss Differences in Afghanistan, Iraq Strategies." *American Forces Information Service*. Retrieved January 10, 2013, from <http://www.centcom.mil/news/petraeus-discusses-differences-in-afghanistan-iraq-strategies>
- Morganthaler, Jeffrey & Giles-Summers, Brandon (2011). *Targeting: Social Network Analysis in Counter IED Operations* (Masters Thesis). Monterey, CA: Naval Postgraduate School. Retrieved January 28, 2013, from <http://hdl.handle.net/10945/5703>
- Muckian, Martin J. (2006). Structural Vulnerabilities of Networked Insurgencies: Adapting Counterinsurgency Strategies to the New Adversary. *Parameters: U.S. Army War College Quarterly*, 36(4), 14-25.
- Nagl, John A. (2005). *Learning to Eat Soup with a Knife: Counterinsurgency Lessons from Malaya and Vietnam*. Chicago, IL: University of Chicago Press.
- Nagl, John A. (2010). Constructing the Legacy of Field Manual 3-24. *Joint Forces Quarterly*, 58, 118-120. Retrieved January 2, 2013, from <http://www.ndu.edu/press/constructing-3-24.html>
- Owen, Mark (2012). *No Easy Day: The Firsthand Account of the Mission that Killed Osama bin Laden*. New York, NY: Dutton.
- Paul, Christopher, & Clarke, Colin P. (2011). Evidentiary Validation of FM 3-24: Counterinsurgency Worldwide, 1978-2008. *Joint Forces Quarterly*, 60, 126-128. Retrieved January 2, 2013, from <http://www.ndu.edu/press/evidentiary-validation-of-FM-3-24.html>
- Petraeus, David H. (2008, September-October). Multi-National Force-Iraq Commander's Counterinsurgency Guidance. *Military Review*, 210-212. Retrieved January 2, 2013, from [http://usacac.army.mil/CAC2/MilitaryReview/Archives/English/MilitaryReview\\_2008CRII0831\\_art028.pdf](http://usacac.army.mil/CAC2/MilitaryReview/Archives/English/MilitaryReview_2008CRII0831_art028.pdf)
- Petraeus, David H. (2010, August 1). COMISAF's Counterinsurgency Guidance. Retrieved January 10, 2013, from [http://www.isaf.nato.int/images/stories/File/COMISAF/COMISAF\\_COIN\\_Guidance\\_01Aug10\\_.doc](http://www.isaf.nato.int/images/stories/File/COMISAF/COMISAF_COIN_Guidance_01Aug10_.doc)
- Petraeus, David H., & Amos, James F. (2006). *Counterinsurgency Field Manual 3-24*. Washington D.C: United States Army. Retrieved January 2, 2013, from <http://www.fas.org/irp/doddir/army/fm3-24.pdf>
- Reed, Brian Joseph (2006). *Formalizing the Informal: A Network Analysis of an Insurgency* (Doctoral dissertation). College Park, MD: University of Maryland.
- Reed, Brian (2007). A Social Network Approach to Understanding an Insurgency. *Parameters*, (Summer):19-30.
- Reed, Brian J., & Segal, David R. (2006). Social Network Analysis and Counterinsurgency Analysis: The Capture of Saddam Hussein. *Sociological Focus*, 39, 251-264.
- Rodriguez, Jose A. Jr. (2013, January 3). A CIA Veteran on What 'Zero Dark Thirty' Gets Wrong about the bin Laden Manhunt. *Washington Post*. Retrieved January 6, 2013, from [http://www.washingtonpost.com/opinions/a-cia-veteran-on-what-zero-dark-thirty-gets-wrong-about-the-bin-laden-manhunt/2013/01/03/4a76f1b8-52cc-11e2-a613-ec8d394535c6\\_story.html?hpid=z2](http://www.washingtonpost.com/opinions/a-cia-veteran-on-what-zero-dark-thirty-gets-wrong-about-the-bin-laden-manhunt/2013/01/03/4a76f1b8-52cc-11e2-a613-ec8d394535c6_story.html?hpid=z2)
- Schmidle, Nicholas (2011, August 8). Getting Bin Laden: What Happened that Night in Abbottabad. *New Yorker*. Retrieved January 1, 2013, from [www.newyorker.com/reporting/2011/08/08/110808fa\\_fact\\_schmidle](http://www.newyorker.com/reporting/2011/08/08/110808fa_fact_schmidle)



- ThinkProgress. (2007). "Breaking: Iraq National Intelligence Estimate Released." Retrieved January 1, 2013, from <http://thinkprogress.org/security/2007/02/02/10034/iraq-nie-released/?mobile=nc>
- Thompson, Robert G.K. (1966). *Defeating Communist Insurgency: The Lessons of Malaya and Vietnam*. London, England: Chatto & Windus.
- Ucko, David (2008, Spring). Innovation or Inertia: The U.S. Military and the Learning of Counterinsurgency. *Orbis*, 52(2), 290-310.
- US Military About (2013). Enlisted Job Descriptions and Qualification Factors. Retrieved July 1, 2013, from [http://usmilitary.about.com/od/enlistedjobs/Enlisted\\_Job\\_Descriptions\\_and\\_Qualification\\_Factors.htm](http://usmilitary.about.com/od/enlistedjobs/Enlisted_Job_Descriptions_and_Qualification_Factors.htm)
- Valentine, Douglas (1990). *The Phoenix Program*. New York, NY: Morrow.
- Woodward, Bob (2008, September 8). Why Did Violence Plummet? It Wasn't Just the Surge. *Washington Post*. Retrieved January 5, 2013, from [http://www.washingtonpost.com/wp-dyn/content/article/2008/09/07/AR2008090701847\\_pf.html](http://www.washingtonpost.com/wp-dyn/content/article/2008/09/07/AR2008090701847_pf.html)
- Woodward, Bob (2009, September 21). McChrystal: More Forces or 'Mission Failure'. *Washington Post*. Retrieved January 17, 2013, from <http://www.washingtonpost.com/wp-dyn/content/article/2009/09/20/AR2009092002920.html>
- Zedong, Mao (1965). A Single Spark Can Start a Prairie Fire. In *Selected Works Vol. 1* (p. 124). Beijing, China: FLP.

## A Computationally Efficient Approximation of Beta Centrality

---

**Zachary Neal**

*Department of Sociology, Michigan State University  
East Lansing, Michigan*

### **Abstract**

Much attention has been focused on developing methods of assessing centrality and power in networks, with particular interest focused on recursive measures that view the status of a node as a function of the status of other nodes. Although such measures have been widely adopted in both academic research and commercial applications, they are computationally intensive and subject to misspecification. This paper introduces a simple measure, alter-based centrality, as a computationally efficient approximation of one commonly used recursive measure, beta centrality. Comparison of these measures in simulated networks indicates that alter-based centrality offers a robust approximation of beta centrality in non-bipartite networks that is computationally efficient and not subject to misspecification.

### **Authors**

*Zachary Neal* is an Assistant Professor of Sociology and Global Urban Studies at Michigan State University. He serves on the editorial boards of *Global Networks* and *City & Community*, as editor of the Routledge “Metropolis and Modern Life” book series, and as Associate Director of the Globalization and World Cities (GaWC) research network. In addition to his work on network methods, his research examines how networks can be used to understand urban phenomena at a range of scales. Additional details may be found on his website at <http://www.msu.edu/~zpneal>.

*Please address correspondence to Zachary Neal, Sociology Department, 316 Berkey Hall, Michigan State University, East Lansing, MI 48824-1111; email: [zpneal@msu.edu](mailto:zpneal@msu.edu).*

## 1. Introduction

Much attention has been focused on developing methods for assessing nodes' centrality within a network, where centrality is often conceptualized as the extent to which a person, organization, web page, or other entity is 'important' for the flow of information or other resources (Freeman, 1979; Borgatti & Everett, 2006). Degree centrality is the simplest approach, focusing only on a node's number of edges. Over several decades, this simple measure has undergone a series of modifications intended to produce more sophisticated measures that look beyond a node's immediate neighbors to assess centrality (Seeley, 1949; Katz, 1953; Hubbell, 1965) and the related construct of power. These incremental revisions have culminated in a recursive approach, which views a node's centrality (or power) as a function of the centrality of the nodes to which it is connected, and to which they are connected, and so on. One recursive measure – beta centrality (Bonacich, 1987), elsewhere known as Bonacich power – has become widely used in academic research (Bearman et al., 2004; Burtis, 2004; Choi et al., 2006; Provan et al., 2009), and has provided the foundation for extensions to commercial applications including Google's PageRank Algorithm (Brin & Page, 1998) and Thompson Scientific's EigenFactor metric for ranking scholarly journals (Bergstrom et al., 2008). However, beta centrality is computationally intensive and subject to misspecification.

In this paper, I examine an alternative measure -- alter-based centrality (Neal, 2011; Neal, 2013)-- that closely approximates beta centrality, but is more computationally efficient and less subject to misspecification errors. Alter-based centrality is not recursive, but instead assesses a node's centrality only within its 2-step neighborhood. Such measures are scattered throughout the literature under various names (Skvoretz & Lovaglia, 1995; Borgatti, 2002; Neal & Neal, 2010; Janoski & Jonas, 2011; Neal & Cappella, 2012), and recently have been adopted in substantive research where some have suggested they may be preferable to beta centrality (Dronen & Lv, 2011). However, to date there has been no formal comparison of these two measures.

## 2. Radial measures of centrality and power

Radial centrality measures can be characterized by the radius within which they assess centrality (Borgatti & Everett, 2006). As a first-order radial measure, degree centrality assesses centrality within a radius of one by simply counting a node's immediate neighbors. Higher-order radial measures of centrality look beyond a node's immediate neighbors, viewing each node's centrality as a function of the centrality of other nodes. These measures focus on quantifying the extent to which a node's position in the network is advantageous in a specific context: diffusion in positively-connected networks. In positively connected networks, such as communication networks where information can diffuse via multiple paths simultaneously, or friendship networks where social status can diffuse via multiple relationships at once, the most important nodes are those that are connected to well connected others (Valente, 1994). Such nodes can diffuse resources like information or

status widely, and can receive resources from a wide range of sources. Indeed, the earliest discussions of recursive measures of centrality sought an index of status (Katz, 1953; Bonacich, 1972), while later discussions focused on communication networks and interpreted the measure's parameters in terms of message transmission probabilities (Bonacich, 1987). The measures of centrality discussed in this paper adopt the general notion that when seeking to diffuse resources widely, or seeking to obtain resources as they diffuse, it is ideal not simply to have many contacts, but to have contacts that are well connected. Thus, they aim not to measure centrality in a universal sense, but rather to measure centrality as the characteristic that makes a node a good point of diffusion or site of concentration.

Higher-order measures of power similarly view each node's power as a function of the centrality of other nodes. However, unlike their centrality counterparts, they focus on quantifying the extent to which a node's position in the network is advantageous in a different context: exchange in negatively-connected networks. In negatively connected networks, such as economic exchange networks where an exchange of capital with one partner reduces the opportunity to exchange with another partner, the most important nodes are those that are connected to poorly connected others. These nodes are able to dominate, or exercise power over, their exchange partners because poorly connected exchange partners have no alternatives and thus are dependent (Emerson, 1962). Adopting this conception of power as dominance in exchange, Cook et al. (1983) sought to understand whether centrality was tantamount to power in exchange relationships. Through a series of experiments and simulations designed to identify which positions in an exchange network allowed actors to generate the most profit, they observed that the most central positions did not obtain the most profit. Thus, they called for "a more general conception of centrality" (p. 298) that could account for this. The measures discussed in this paper respond to this call by aiming to quantify the notion that when seeking to dominate an exchange relationship, it is ideal not simply to have many contacts, but to have contacts that can be dominated because they are poorly connected and lack alternative exchange partners. Thus, they aim not to measure power in a universal sense, but rather to measure power as the characteristic that gives a node the 'upper hand' in exchange.

### 2.1 Beta centrality

Driven by the logic that "if second order indices are better than first order indices, then third order indices...should be even better," beta centrality assesses centrality within an infinite radius, allowing each node's centrality to contribute to the centrality of other nodes (Bonacich, 1972, p. 114). Thus, it can be expressed as an infinite summation of a node's alters' degree centralities as

$$BC = \sum_{k=0}^{\infty} \beta^k \mathbf{A}^{k-1} \mathbf{1} = \mathbf{A} \mathbf{1} + \beta \mathbf{A}^2 \mathbf{1} + \beta^2 \mathbf{A}^3 \mathbf{1} \dots \quad (1)$$

where  $\mathbf{A}$  is an adjacency matrix and  $\mathbf{1}$  is a column vector of 1s. The  $\beta$  parameter attenuates the weight given to neighbors at successively greater distances. This infinite sum converges on a unique set of values when  $|\beta| < 1/\lambda_1$ , where  $\lambda_1$  is the largest eigenvalue of  $\mathbf{A}$ . When this condition is met

$$BC = (\mathbf{I} - \beta\mathbf{A})^{-1} \mathbf{A}\mathbf{1} \quad (2)$$

where  $\mathbf{I}$  is an identity matrix (Bonacich, 1987). Although equation (2) avoids the intractability of an infinite sum, it does so at the cost of a computationally intensive matrix inversion.

The sign of  $\beta$  determines the substantive interpretation of  $BC$ . When  $\beta > 0$ ,  $BC$  is a measure of centrality, assigning higher scores to nodes that are connected to more central others. In contrast, when  $\beta < 0$ ,  $BC$  is a measure of power, assigning higher scores to nodes that are connected to less central others.<sup>1</sup> Finally, when  $\beta = 0$ ,  $BC$  is proportional to degree centrality. Throughout this paper, I distinguish the centrality and power forms of beta centrality using  $BC^+$  and  $BC^-$ , respectively.

The magnitude of  $\beta$  determines the extent to which successively more distant contacts contribute to a node's centrality or power. The selection of an appropriate value for  $\beta$  is important because  $BC$  "can give radically different rankings on centrality, depending on the value of  $\beta$ " (Bonacich, 1987), and because use of values greater (in absolute value) than  $1/\lambda_1$  yield meaningless scores. In practice, however, misspecification errors are common, with  $\beta$  set either too low (Choi et al., 2006), or too high (Irwin & Hughes, 1992; Hanneman & Riddle, 2005; Weir et al., 2009). In the analyses presented below,  $BC^+$  is computed by setting  $\beta = 0.99/\lambda_1$  and  $BC^-$  is computed by setting  $\beta = -0.99/\lambda_1$ . Using the highest and lowest allowable values for  $\beta$  maximizes  $BC$ 's ostensibly key advantage over ordinary degree centrality: to assess centrality and power by taking into account the network's global structure.

### 2.2 Alter-based centrality

Alter-based centrality challenges  $BC$ 's motivating assumption that third- and higher-order indices are better than second-order indices. Thus, alter-based centrality assesses nodes' centrality only within a 2-step radius. Specifically, alter-based centrality is the sum of a node's alters' degree centralities. Because well-connected friends are better sources of information than poorly connected friends, this views an alter's contribution to ego's centrality as proportional to the alter's centrality. Alter-based centrality can be computed as

$$AC^+ = \mathbf{A}(\mathbf{A}\mathbf{1}) \quad (3)$$

A small modification to equation (3) yields a complementary measure power

$$AC^- = \mathbf{A} \left( \frac{1}{\mathbf{A}\mathbf{1}} \right) \quad (4)$$

where the division is performed element-wise. When used to measure power, alter-based centrality is the sum of a node's alters' inverse degree centralities. Because poorly connected friends are easier to control than well connected friends, this views an alter's contribution to ego's power as inversely proportional to the alter's centrality. As reflected in equations (3) and (4), throughout this paper I distinguish the centrality and power forms of alter-based centrality using  $AC^+$  and  $AC^-$ , respectively.

There are two important differences between  $BC$  and  $AC$ . First, the computation of  $AC$  does not involve the selection of an attenuation parameter ( $\beta$ ), and thus is not subject to misspecification errors. Second, the computation of exact values of  $AC$  does not require a computationally costly matrix inversion.

### 3. Comparison of beta centrality and alter-based centrality

#### 3.1 Illustrative example

Figure 1 compares the relative  $BC$  and  $AC$  scores assigned to different positions in a simple, illustrative network. When used to assess centrality, both  $BC^+$  and  $AC^+$  assign the highest scores to the center, black node. This highlights that this node is connected to several others (i.e. the gray nodes), that themselves are well-connected. They assign somewhat lower scores to the gray nodes, highlighting that while these positions are also well connected, they are connected mainly to poorly connected others (i.e. the white nodes). When used to assess power, a different picture emerges.  $BC^-$  and  $AC^-$  assign the highest score to the gray nodes, reflecting their connection to poorly connected others; in exchange situations, the grey nodes can dominate (i.e. dictate the terms of exchange with) the white nodes. They assign lower scores to the black node, reflecting

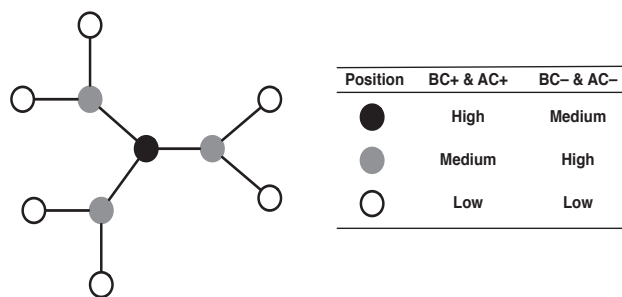


Figure 1. Beta centrality and alter-based centrality in a hypothetical network.

<sup>1</sup> Strictly speaking, when  $\beta < 0$ , negative weight is given to the centrality of nodes that are an odd number of steps away (i.e. neighbors' centrality reduces one's power), but positive weight is given to the centrality of nodes that are an even number of steps away (i.e. neighbors of neighbors' centrality increases one's power). This feature can be seen in equation (2), where  $\beta$  is raised to successively higher powers when applied to more distant nodes, such that when  $\beta < 0$ , it is negative when raised to an odd power but is positive when raised to an even power.

the fact that while this node is well connected, it is connected to others (i.e. the gray nodes) who have alternative exchange partners and thus cannot be dominated. Finally, whether used to assess centrality or power, *BC* and *AC* both assign the lowest scores to the white nodes, reflecting their non-central and powerless status as pendants.

### 3.2. Computational efficiency

Figure 2 compares the computer running time necessary to compute exact *BC* and *AC* scores in sparse (density = 0.05) networks ranging in size from 10 to 10,000 nodes. For small networks, both measures can be computed quickly and in roughly similar times, making the benefits of computational efficiency offered by *AC* negligible in such cases. However, for larger networks, exact values of *BC* take much longer to compute. This is important because attention has increasingly turned to the analysis of very large networks, such as actor collaborations ( $N = 225,226$ ; (Watts & Strogatz, 1998)) and the world wide web ( $N = 325,729$ ; (Albert et al., 2000)). In these cases, such an improvement in computational efficiency is quite dramatic, and indeed, the matrix inversion necessary to compute *BC* in networks of this size would be impossible using typically available computing resources.

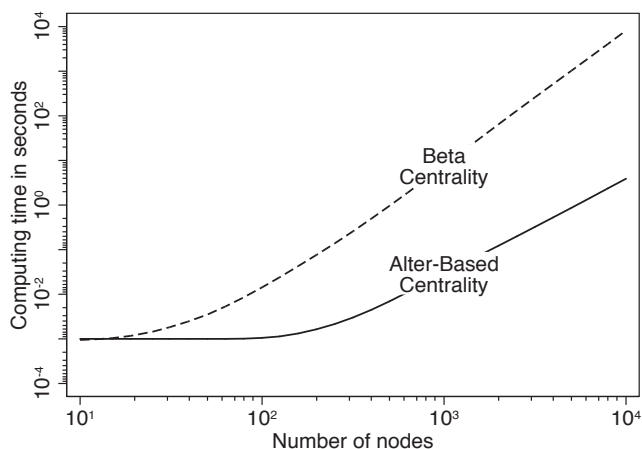


Figure 2. Comparison of computational efficiency.

### 3.3 Similarity in simulated networks

To examine *AC*'s potential as an approximation of *BC*, I examined the Pearson correlation coefficient between *AC+* and *BC+* (*rCentrality*), and between *AC-* and *BC-* (*rPower*), in 1,000,000 Erdős-Rényi graphs (Bolland, 1988; Valente et al., 2008; Barnett, 2010). All simulated networks were symmetric, binary, and structurally unique (i.e. structurally isomorphic replicates were excluded). Regular and maximally centralized networks were excluded as trivial cases.<sup>2</sup> This procedure sam-

ples randomly from the set of all possible networks containing between 5 and 50 nodes. Table 1 summarizes several key structural characteristics of the random networks, and highlights that the sample of simulated networks examined below includes those with a broad range of structural characteristics. For example, it includes both dense and sparse networks, those with symmetric and skewed degree distributions, and those exhibiting both assortativity and disassortativity.

Table 1. Characteristics of simulated networks ( $N = 1,000,000$ )

Structural Characteristic	Mean	S.D.	Minimum	Maximum
Size	30.75	11.85	5	50
Density	0.55	0.24	0.05	0.99
Degree distribution ( $\sigma/\mu$ )	0.19	0.12	0.01	1.00
Degree distribution (skew)	-0.05	0.48	-4.50	2.77
Degree assortativity	-0.08	0.09	-0.93	0.72
Spectral bipartivity	0.54	0.09	0.50	1.00
Core-Peripherality ( $\lambda_1/\lambda_2$ )	5.78	6.23	1.03	118.85

The centrality scores yielded by *AC+* and *BC+* are very highly correlated in these networks (*rCentrality* mean = 0.997, *sd* = 0.007). Similarly, the power scores yielded by *AC-* and *BC-* are also highly correlated (*rPower* mean = 0.988, *sd* = 0.033). Figure 3 illustrates the heavily skewed distribution of *rCentrality* (left panel) and *rPower* (right panel); *rCentrality* < .75 in only 27 (0.003%) of the simulated networks, while *rPower* < .75 in 2657 (0.266%) simulated networks. These results indicate that *AC* and *BC* yield nearly identical scores under most circumstances. This close relationship is striking because, although *BC* is intended to consider the network's global structure when quantifying a node's position, it yields the same scores as *AC*, which is explicitly restricted to assessing status within a node's 2-step neighborhood. Thus, these results call into question whether *BC* can truly be interpreted as assessing centrality within an infinite radius.

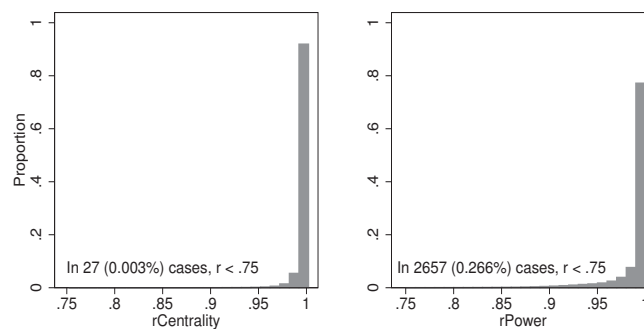


Figure 3. Correlations of centrality and power scores in simulated networks.

<sup>2</sup>Notably, *AC+* does not conform to conventional notions of centrality when applied to maximally centralized networks like stars, where it assigns all nodes a score of  $N - 1$ . This reflects the ambiguity in status between two possible situations: (a) a connection to a single, highly connected alter, and (b) connections to many, poorly connected alters. Assessing node centrality in a star network is a trivial exercise, but for those who desire a measure that will assign the highest value to the center node, *AC+* would not be appropriate.

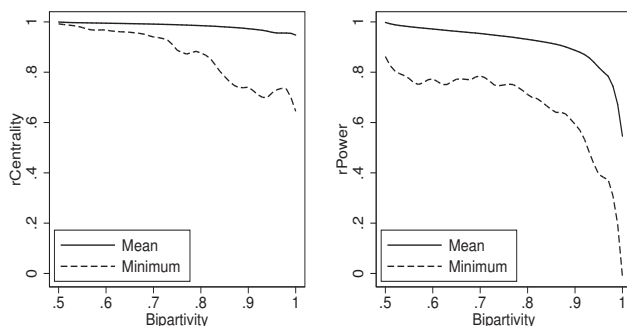
These high correlations suggest that *AC* is a computationally efficient approximation of *BC*. However, in a small number of cases, these two measures yield different scores, raising questions about when *AC* is not an approximation of *BC*. To investigate this, I regressed the structural characteristics shown in Table 1 on *rCentrality* and *rPower*. The standardized regression coefficients shown in Table 2, which are all significant at  $p < .001$  using heteroskedasticity robust standard errors, were obtained using ordinary least squares regression, which is appropriate in this context because each simulated network represents an independent observation. These results indicate that spectral bipartivity plays the greatest role in determining when *AC* approximates *BC*, with higher levels of bipartivity leading to greater divergence. Figure 4 displays the mean and minimum values of *rCentrality* and *rPower* observed in simulated networks with varying levels of bipartivity, illustrating that *AC* and *BC* yield nearly identical scores when applied to non-bipartite networks, and differing scores only when applied to highly bipartite networks. More formally (Estrada et al., 2005):

$$\lim_{x \rightarrow 0.5} AC = BC, \text{ where } x = \frac{\sum_{j=1}^N \cosh(\lambda_j)}{\sum_{j=1}^N e^{\lambda_j}} \quad (5)$$

**Table 2.** Effect of structural characteristics on the correlation.

Structural Characteristic	rCentrality	rPower
Size	-0.023	0.027
Density	0.326	0.080
Degree distribution ( $\sigma/\mu$ )	0.358	-0.022
Degree distribution (skew)	-0.043	0.033
Degree assortativity	0.228	-0.062
<b>Spectral bipartivity</b>	<b>-0.595</b>	<b>-0.541</b>
Number of pendants	-0.219	-0.264
Core-Peripherality ( $\lambda_1/\lambda_2$ )	-0.039	-0.016
$R^2$	0.578	0.628

Note: Standardized regression coefficients; all are significant at  $p < .001$  using heteroskedasticity robust standard errors.



**Figure 4.** Mean and minimum correlations, by level of bipartivity.

#### 4. Discussion

The greater computational efficiency of *AC* compared to *BC* is relatively unsurprising given the former measure’s simplicity. However, this simplicity is deceptive because, as figures 3 and 4 illustrate, *AC* closely approximates *BC* scores in non-bipartite networks. Notably, *AC+* remains a reasonable approximation of *BC+* even in bipartite networks, but *AC-* fails to approximate *BC-* under these conditions. However, this restriction may not be practically significant because highly bipartite networks are rare in social and information contexts, with a network of sexual relations among heterosexual individuals standing as a notable exception (Estrada et al., 2005).

Nonetheless, the question remains, why do *AC* and *BC* differ in bipartite networks? Additional investigation using experimental or simulated exchange networks will be necessary to answer this question. However, one possibility is that higher-order radial measures are simply not appropriate in bipartite networks. In a bipartite network, edges exist only between nodes partitionable into two mutually exclusive sets. All even-length walks terminate at a node in the same set, while all odd-length walks terminate at a node in a different set. Thus, first-order (e.g. degree centrality) measures consider only the impact of nodes in a different set, with which exchanges of resources are possible. In contrast, higher- (e.g. alter-based centrality) and infinite-order (e.g. beta centrality) measures have a much wider scope, and also consider the impact of nodes in the same set, with which exchanges of resources are structurally impossible in a bipartite network. In essence, when applied to a bipartite network, higher-order measures simply ‘look too deep’ into the network’s structure. Furthermore, a unique feature of *BC-*’s computation may help explain why it diverges from *AC-* more dramatically than *BC+* does from *AC+*. Recall that *BC-* positively weights the centrality of alters located an even number of steps from ego, but negatively weights the centrality of alters located an even number of steps away. This has an unusual consequence when applied to a bipartite network. Specifically, in a bipartite network, *BC-* positively weights the centrality of alters in ego’s own set and negatively weights the centrality of alters in the other set. In natively bipartite (i.e. two-mode) network data, the assessment of centrality using measures designed specifically for such a context (see Borgatti & Everett, 1997) is likely more appropriate than using either *BC* and *AC*, which are primarily intended for the analysis of traditional one-mode networks.

In non-bipartite networks *AC* yields close approximations of the more familiar but more complex *BC*. However, there are two additional reasons that may recommend the use of *AC* over *BC*: its applicability to multi-component networks and its minimal data requirements. First, *BC* is known to yield correct scores only for the largest component in a network containing multiple components (Bonacich, 2007). In such cases, *BC* must be computed for each component separately, and each component’s vector of scores must be normalized before comparison to those of the other components. Thus, in a multi-component network, computation of *BC* involves at least three steps: (1) extraction of each component, (2) computation of *BC* within each component, each time requiring a separate

matrix inversion, and (3) application of a normalization. Not only does this add additional steps to the analysis, but it also ignores the fact that there is an important distinction between (a) a single network composed of multiple components and (b) two fundamentally different networks. For example, analyzing children's relationships in a single classroom where boys and girls happen to (but did not necessarily need to) form two disconnected components (i.e. cliques) is not the same as analyzing two entirely different networks (e.g. a classroom friendship network and a network of internet routers) where the absence of edges between children and routers is necessary. In contrast, *AC* can be directly computed in multi-component networks, thus preserving the network's multi-component nature (i.e. that boys and girls could have been connected, even if they weren't), and will yield the same scores whether computed on the network as a whole or on each component individually.

Second, the computation of *BC* requires whole-network data, that is, information on all relationships among the actors in a given setting. As a result, those with ego-centric network data have not been able to use it. In contrast, *AC* only requires data on an actor's ego-network and the degree of each alter. Thus, *AC* allows those with ego-centric network data to compute something very closely resembling *BC*, which previously was not possible. That is, *AC* reduces the data requirements for computing radial point-centrality indices.

Future research may explore the role of bipartivity on radial measures of centrality, and on the extension of *AC* to the context of valued and directed networks. However, these initial analyses suggest that alter-based centrality offers a useful alternative to beta centrality. Beta centrality remains a flexible radial measure because, through modification of the  $\beta$  parameter, it can measure both centrality (when  $\beta > 0$ ) in positively connected networks and power (when  $\beta < 0$ ) in negatively connected networks. However, it is computationally complex, prone to misspecification, and restricted to connected components and whole-network data. Despite its simplicity, *AC* offers a computationally efficient approximation of *BC* in non-bipartite networks and remains a reasonable approximation of *BC+* in bipartite networks. Additionally, it is not subject to misspecification and can be applied to multi-component networks and ego-centric network data.

## References

- Albert, R., Jeong, H., & Barabási, A.-L. (2000). Error and attack tolerance of complex networks. *Nature*, *406*, 378 – 382.
- Barnett, G. A. (2010). A measure of centrality for dense networks with valued ties. *Connections*, *30*, 11 – 20.
- Bearman, P.S., Moody, J., & Stovel, K. (2004). Chains of affection: adolescent romance and sexual networks. *American Journal of Sociology*, *110*, 44 – 91.
- Bergstrom, C. T., West, J. D., & Wiseman, M. A. (2008). The eigenfactor metrics. *Journal of Neuroscience*, *28*, 11433 – 34.
- Bolland, J. M. (1988). Sorting out centrality: an analysis of the performance of four centrality models in real and simulated networks. *Social Networks*, *10*, 233 – 253.
- Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, *2*, 113 – 20.
- Bonacich, P. (1987). Power and centrality: a family of measures. *American Journal of Sociology*, *92*, 1170 – 82.
- Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, *29*, 555 – 564.
- Borgatti, S.P., & Everett, M.G. (1997). Network analysis of 2-mode data. *Social Networks*, *19*, 243 – 269.
- Borgatti, S.P., & Everett, M.G. (2006). A graph-theoretic perspective on centrality. *Social Networks*, *28*, 466 – 84.
- Borgatti, S.P. (2002). NetDraw: Graph Visualization Software, Harvard: Analytic Technologies.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, *30*, 107 – 17.
- Burris, V. (2004). The academic caste system: prestige hierarchies in PhD exchange networks. *American Sociological Review*, *69*, 239 – 64.
- Choi, J.H., Barnett, G.A., & Chon, B. (2005). Comparing world city networks: a network analysis of internet backbone and air transport intercity linkages. *Global Networks*, *6*, 81 – 99.
- Cook, K.S., Emerson, R.M., Gillmore, M.R., & Yamagishi, T. (1983). The distribution of power in exchange networks: theory and experimental results. *American Journal of Sociology*, *89*, 275 – 305.
- Dronen, N., & Lv, Q. (2011). Return probability and k-step measures, arXiv:1105.4479v1 [physics.soc-ph].
- Emerson, R. M. (1962). Power-dependence relations. *American Sociological Review*, *27*, 31 – 41.
- Estrada, E., & Rodríguez-Velázquez, J. A. (2005). Spectral measures of bipartivity in complex networks. *Physical Review E*, *72*, 046105.
- Freeman, L.C.. (1979). Centrality in networks I. conceptual clarification. *Social Networks*, *1*, 215 – 39.
- Hanneman, R., & Riddle, M. (2002). *Introduction to Social Network Methods*. Riverside, CA: University of California, Riverside.
- Hubbell, C.H. (1965). An input-output approach to clique identification. *Sociometry*, *28*, 377 – 99.
- Irwin, M. D., Hughes, H. L. (1992). Centrality and the structure of urban interaction: measures, concepts, and applications. *Social Forces*, *71*, 17 – 51.
- Janoski, T., Jonas, A. (2011, August 20). A synthetic theory of political sociology: bridging social networks and power dependence to power resources theory in city politics. Presented at the American Sociological Association, Las Vegas, NV.
- Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, *18*, 39 – 43.
- Neal, J.W., Neal, Z.P. (2010). Power as a structural phenomenon. *American Journal of Community Psychology*, *48*, 157 – 167.
- Neal, Z.P. (2011). Differentiating centrality and power in the world city network. *Urban Studies*, *48*, 2735 – 2750.

- Neal, J.W., & Cappella, E. (2012). An examination of network position and childhood relational aggression: integrating resource control and social exchange theories. *Aggressive Behavior*, 38, 126 – 140.
- Neal, Z.P. (2013). Does world city network research need eigenvectors? *Urban Studies*, 50, 1648 – 1659.
- Provan, K. G., Huang, K., & Milward, H. B. (2009). The evolution of structural embeddedness and organizational social outcomes in a centrally governed health and human services network. *Journal of Public Administration Research and Theory*, 19, 838 – 93.
- Seeley, J.R. (1949). The net of reciprocal influence: a problem in treating sociometric data. *Canadian Journal of Psychology*, 4, 234 – 40.
- Skvoretz, J., Lovaglia, M. J. (1995). Who exchanges with whom: structural determinants of exchange frequency in negotiated exchange networks. *Social Psychology Quarterly*, 58, 163 – 77.
- Valente, T. W. (1994). *Network Models of the Diffusion of Innovations*. Cresskill, NJ: Hampton.
- Valente, T. W., Coronges, K., Lakon, C., & Costenbader, E. (2008). How correlated are network centrality measures. *Connections*, 28, 16 – 26.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393, 440 – 442.
- Weir, M., Rongerude, J., & Ansell, C. K. (2009). Collaboration is not enough: virtuous cycles of reform in transportation policy. *Urban Affairs Review*, 44, 455 – 489.



# Community Structure in Multi-Mode Networks: Applying an Eigenspectrum Approach

---

**David Melamed**

*University of South Carolina*

**Ronald L. Breiger**

*University of Arizona*

**A. Joseph West**

*University of Arizona*

## Abstract

We combine the logic of multi-mode networks developed in Fararo and Doreian (1984) with Newman's (2006) spectral partitioning of graphs into communities. The resulting generalization of spectral partitioning provides a simple, elegant, and useful tool for discovering the community structure of multi-mode graphs. We apply the generalized procedure to a published three-mode network and find that the results of the algorithm are consistent with existing substantive knowledge. We also report the results of extensive simulations, which reveal that the generalization becomes more effective as the networks become denser.

## Authors

*David Melamed* is currently an Assistant Professor of Sociology at the University of South Carolina. His research interests are in group processes, social networks, mathematical modeling, and stratification.

*Ronald L. Breiger* is Professor of Sociology at the University of Arizona and holds affiliate appointments in the Graduate Interdisciplinary Program in Statistics and in the School of Government and Public Policy. His research interests include social network modeling, culture, and adversarial networks.

*A. Joseph West* is a PhD Candidate in Sociology at the University of Arizona. His research interests lie at the intersection of social networks, culture, and social movements.

## Notes

This research was supported by grants from the Defense Threat Reduction Agency (HDTRA1-10-1-0017) and the Air Force Office of Scientific Research (FA9550-10-1-0569) for which the second-listed author is the PI and a Co-PI, respectively. We thank Sean Everton and two anonymous reviewers for helpful feedback on this project.

*Please address correspondence to David Melamed, Department of Sociology, University of South Carolina, Sloan College 321, 911 Pickens Street, Columbia, SC 29208. Email: melamed@sc.edu*

## 1. Introduction

Discovering cohesive subgroups, cliques, modules, or communities in networks has a rich history in the computer and social sciences (Felder, 1973; Pothén, Simon & Liou, 1990; Wasserman & Faust, 1994; Jackson, 2008), but has seen an explosion of development since Girvan and Newman (2002) brought this problem to the attention of the mathematics and statistical physics communities (Porter, Onnela & Mucha, 2009). Communities within networks refer to densely connected subsets of vertices or nodes within the network. Several approaches have been leveraged to optimize solutions to identify communities, including partitioning clustering procedures (e.g., Porter, Mucha, Newman, & Friend, 2007), centrality-based procedures (e.g., Girvan & Newman, 2002), and k-clique-based procedures (e.g., Palla, Derenyi, Farkas & Vicsek, 2005), among others. We focus on Newman's (2006) spectral partitioning approach to modularity maximization because it is simple, intuitive, and quite popular.

Specifically, we generalize spectral partitioning to multi-mode networks, or networks that consist of more than one type of vertex (e.g., persons, groups, and events), using the logic for multi-mode networks set forth by Fararo and Doreain (1984). On the one hand, there exist several applications of modularity-based community algorithms to two-mode networks (e.g., Zhang et al., 2008). However, these applications rely on projections linking vertices in one mode (e.g., people) to other vertices in that mode through their mutual relations in the second mode (e.g., committees) (see Breiger, 1974). That is, some applications transform two-mode networks of who-to-what into one-mode networks of who-to-whom, and then apply community finding algorithms to the projections. On the other hand, Barber (2007) generalized modularity-based methods for partitioning networks into communities to the realm of bipartate graphs or two-mode networks, and Guimerà, Sales-Pardo, and Amaral (2007) developed an algorithm for examining the community structure of one of the modes in a bipartite graph. Here, we generalize modularity-based methods to n-mode networks, and apply the resulting algorithm to a published example, illustrating the utility and accuracy of the procedure based on substantive and qualitative knowledge. We also apply the algorithm to large simulated networks, illustrating some properties associated with its efficiency.

Community detection in networks or graphs seeks to partition the vertices into communities within which there is a concentration of ties. In social networks there may be cliques or groups of friends within which many ties are shared, while relatively few ties are sent to the rest of the network. On the internet there are communities of web sites with related topics that share links at above average rates, and yet send few links outside their community. The task, then, of community detection algorithms is to determine a useful way to partition the network into communities. One metric that has been developed in this connection is modularity, which reflects the extent, relative to a null model, to which edges are found within communities instead of between communities. Modularity provides a benchmark for comparing possible partitions of the vertices in a network.

Unfortunately, there is no way to ensure that any modularity solution is the optimal solution; optimization of the community structure is known to be NP-hard, and several procedures, old and new, have been leveraged with respect to optimization (Porter et al., 2009). The procedure we focus on here for optimizing modularity is based on spectral partitioning of the so-called "modularity" matrix. Specifically, we use the eigenspectrum of the modularity matrix, which we adjust to account for multi-mode networks. Below we review modularity maximization based on the eigenspectrum. We next elaborate an appropriate means of applying the maximization to multi-mode networks, and we apply the algorithm to a published three-mode network. We then report the results from simulations of large four-mode networks that illustrate the utility and flexibility of our approach.

## 2. Modularity Maximization

The Newman algorithm based on the eigenspectrum of a network is elegantly simple. Consider a network with  $n$  vertices (nodes) and  $m$  edges (relations) defined by a binary symmetric adjacency matrix  $\mathbf{A}$  where an edge is denoted by a '1' and  $\mathbf{A}_{ij}$  is '0' otherwise. Let  $\mathbf{P}$  denote a matrix under a null model such that the  $\mathbf{P}_{ij}$  are probabilities in the null model that an edge exists between vertices  $i$  and  $j$ . In this paper, and in most cases, the null model is a simple model of independence such that  $P_{ij} = P_{i+} P_{+j}$ . Now the so-called modularity matrix ( $\mathbf{B}$ ) refers to the difference between  $\mathbf{A}$  and  $\mathbf{P}$ :

$$B_{ij} = A_{ij} - P_{ij} \quad (1)$$

Newman (2006) recommends using the eigenspectrum of  $\mathbf{B}$  to partition the vertices into modules. Specifically the eigenvector associated with the leading eigenvalue of  $\mathbf{B}$  partitions the vertices into an optimal two-community solution such that the vertices in one community will have positive (or zero) eigenvector scores and those in the other community will have negative scores (Newman, 2006). Subsequent splits of the vertices into more than two communities can be identified similarly by looking at the signs of the eigenvector associated with the second leading eigenvalue, and so on. In general, the upper bound on the number of communities that may be found in this way is equal to one plus the number of positive eigenvalues of  $\mathbf{B}$  (Newman, 2006). The algorithm uses these splits to identify communities, but the preferred solution is the one that maximizes the modularity  $Q$ . That is, the algorithm converges when any subsequent split of the community structure makes a zero or negative contribution to the modularity  $Q$ .

Let  $c$  denote the number of modules or communities that a graph is to be partitioned into. Then let  $\mathbf{S}$  denote an  $n \times c$  matrix where each row is an index vector indicating vertex  $i$ 's membership in module  $j$  by a '1' with '0's elsewhere in the row. Each node may be assigned to only one module, making the unit vectors orthogonal (Barber, 2007). With  $\mathbf{S}$  thusly defined, the modularity is as follows:

$$Q = 1 / (2m) \times Tr \mathbf{S}^T \mathbf{B} \mathbf{S} \quad (2)$$

Where  $Tr$  refers to the trace of the matrix product  $\mathbf{S}^T\mathbf{B}\mathbf{S}$  and  $\mathbf{S}^T$  refers to the transpose of matrix  $\mathbf{S}$  (Barber, 2007; Newman, 2006).<sup>1</sup> Maximization of  $Q$  yields the optimal community structure of the graph.<sup>2</sup>

### 3. Multi-Mode Networks and Modularity

In addition to one-mode networks, bipartite networks have been a focus of some analyses of community structure (Barber, 2007; Zhang et al., 2008), but tripartite or quadripartite networks have received much less attention (c.f. Mucha et al., 2010; Murata, 2011, for exceptions). Bipartite networks have a much simpler structure than do generalized multi-mode networks because one mode constitutes the rows of a matrix while the other mode constitutes the columns. With three or more modes, adjustments are required to place the modes in the same adjacency space. Fararo and Doreain (1984; see also Carley, 2003) have shown how to compile multi-mode networks where the number of modes exceeds two. We review this strategy and discuss how to transform these multi-mode networks into modularity matrices amenable to spectral partitioning.

Three-mode networks (or  $n+2$ -mode networks, where  $n$  is any positive integer) require a block off-diagonal form in order to put all of the modes into a single adjacency space. Of course, this assumes there are no within-mode ties. For example, assume a network of edges between three types of vertices: persons, committees, and organizations. Denote the matrix of membership of persons on committees by  $\mathbf{C}$ , the matrix of persons' affiliation with organizations by  $\mathbf{D}$ , and the matrix linking committees to organizations by  $\mathbf{E}$ . (We assume that some committees may draw members from multiple organizations, and that similar cross-cutting affiliations are possible with respect to all pairs of modes.) Then the block off-diagonal matrix representation of the three-mode network, denoted  $\mathbf{Z}$ , may be represented as follows (Fararo & Doreian, 1984):

$$\mathbf{Z} = \begin{pmatrix} 0 & \mathbf{C} & \mathbf{D} \\ \mathbf{C}^T & 0 & \mathbf{E} \\ \mathbf{D}^T & \mathbf{E}^T & 0 \end{pmatrix} \quad (3)$$

Matrix  $\mathbf{Z}$  may not be directly transformed into a modularity matrix ( $\mathbf{B}$ ) because that would violate that matrix  $\mathbf{Z}$  be block off-diagonal. Consequently, we propose to compute the null model separately for each of the two-mode matrices that constitutes the three-mode matrix, to subtract out the null from each two-mode matrix, and then to aggregate them into the full three-mode modularity matrix. Let  $\mathbf{C}_A$ ,  $\mathbf{D}_A$ , and  $\mathbf{E}_A$  denote the two-mode adjacency matrices, let  $\mathbf{C}_B$ ,  $\mathbf{D}_B$ , and  $\mathbf{E}_B$  denote the

null models for matrices  $\mathbf{C}$ ,  $\mathbf{D}$ , and  $\mathbf{E}$ , and let  $\mathbf{C}_B$ ,  $\mathbf{D}_B$ , and  $\mathbf{E}_B$  denote the modularity matrices for  $\mathbf{C}$ ,  $\mathbf{D}$ , and  $\mathbf{E}$  (i.e.,  $\mathbf{C}_B = \mathbf{C}_A - \mathbf{C}_p$ ,  $\mathbf{D}_B = \mathbf{D}_A - \mathbf{D}_p$ ,  $\mathbf{E}_B = \mathbf{E}_A - \mathbf{E}_p$ ). Then the appropriate three-mode adjacency matrix on which to compute the modularity  $Q$  is of the following form:

$$\mathbf{Z}_B = \begin{pmatrix} 0 & \mathbf{C}_B & \mathbf{D}_B \\ \mathbf{C}_B^T & 0 & \mathbf{E}_B \\ \mathbf{D}_B^T & \mathbf{E}_B^T & 0 \end{pmatrix} \quad (4)$$

Although  $\mathbf{Z}_B$  is a three-mode matrix, the logic of its compilation can be extended to any number of modes. West, Melamed, and Breiger (2012) have applied this procedure to finding communities in four-mode narrative networks of people, groups, events, and games. We now turn to an example using a published three-mode network.

### 4. Ecology of Games and Tripartite Networks

Cornwell, Curry and Schwirian (henceforth CCS; 2003) analyzed a three-mode network consisting of actors, issues, and games organized around a major conflict in an urban community: the construction of a large-scale sports stadium in Cincinnati, OH during the 1990's. The three "modes" included five actors (some of whom were individuals, such as the general manager of Cincinnati's football team, and some organizational, such as the City Council), nine issues (for example, creating a referendum on whether to build a new stadium), and six games (such as the territorial game and the sports franchise game).<sup>3</sup> As the total number of nodes was only 20, the multi-mode network data collected by CCS provides an ideal didactic example for demonstrating the usefulness of our proposed procedure. CCS aimed to implement network techniques and procedures of analysis to formalize Long's (1958) ecology of games perspective, where games refer broadly to agendas and the domains within which they are pursued. CCS used multi-dimensional scaling and variants of network density to identify key nodes and groupings of nodes, and argued that their approach to the ecology of games aids in understanding the structure and process of community affairs.

CCS also published their data (2003, p. 133), along with much qualitative information about the controversy and the networks that were implicated in it. The key players (actors) in this network are Mike Brown, the manager of the Cincinnati Bengals football team, Marge Schott, the owner of the Cincinnati Reds baseball team, the City Council, the County Commissioners, and the general public. The issues consist of the new facility (NewFacility) that would keep the Bengals in

<sup>1</sup> Barber (2007) uses  $1/2m$  in eq. 2, while Newman (2006) uses  $1/4m$ . The choice is arbitrary since for any possible community structure  $m$  is fixed. Newman justifies his choice only with respect to comparability to earlier formulations of modularity

<sup>2</sup> Vertices partitioned into separate modules on one split cannot be reallocated into the same module on subsequent splits, and hence the solution may not be "optimal." This is an example of why a technically optimal solution is NP-hard (see also, Brandes et al., 2008).

<sup>3</sup> A main substantive goal of Cornwell et al. (2003) was to demonstrate the effectiveness of multi-mode network modeling in providing an analytic framework for applying the approach to studying the local urban community as "an ecology of games" that was pioneered by Norton Long (1958).

Cincinnati, proposed renovations to the old stadium (ProposingStadiumRenovations), the creation of the referendum (CreatingReferendum) to build a new stadium, passing the referendum (PassingReferendum), where to build the new stadium (WhereToBuild), determining the preliminary site (PrelimSite), actually drafting the terms of the deal (DraftingTerms), stalling the transfer of the land (LandTransfer), and reaching a transfer agreement on the land (ReachingAgreement). The games consist of politics (Politics), urban redevelopment (UrbanDevelopment), sports franchises (SportsFranchise), business competition (BusinessCompetition), territory (Territory), and the budget (Budget). For more detail on any of these nodes, please see CCS (2003, pp. 128-132). The three-mode network is visually represented in Figure 1 with triangles denoting people, circles denoting issues, and squares denoting games.

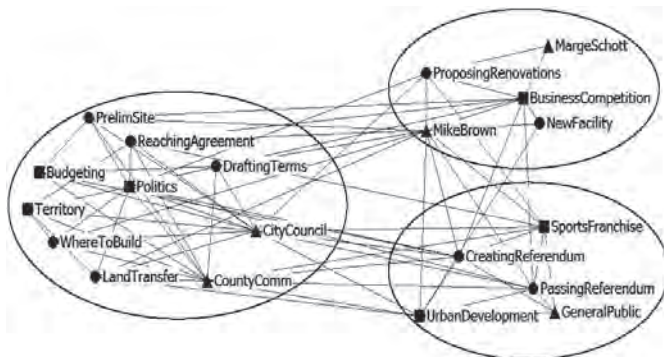


Figure 1. Sociogram of three-mode network. Ellipses denote communities and shapes denote modes (triangles are persons, circles are issues, and squares are games).

To determine the community structure of the CCS network, we computed the modularity matrices for each of the two-mode matrices that constitute the full three mode matrix (i.e., the  $C_B$ ,  $D_B$ , and  $E_B$  matrices from above), and computed its eigenstructure.<sup>4</sup> The results of the algorithm suggest that a three-community solution results in the maximum modularity ( $Q \approx .051$ ). Specifically, nodes that are positive on the leading eigenvector form one community, nodes that are negative on the leading eigenvector and positive on the second leading eigenvector form a second community, and nodes that are negative on both the first and second leading eigenvectors form a third community. The results of the algorithm are also presented in Figure 1, with communities denoted by ellipses. Figure 1 also shows that our optimal three-community solution combines actors, issues, and games within each of the identified communities.

CCS (2003, p. 135) present the results of a multidimensional scaling of a distance matrix derived from the three-mode network. Based on their substantive and qualitative knowledge of the network, they indicated two communities in the network, leaving a few nodes out of either community. The largest community they identify contains five issues (where to build, the preliminary site, drafting of the terms, the land transfer, and reaching an agreement), two players (the county commissioners and the city council) and three games (politics, the budget, and territory). Here we point out that our eigenspectrum ap-

proach to community finding identified *the exact same community* without any substantive knowledge. This is the community to the left in Figure 1. This community can generally be thought of as the community that formed around the logistics of building the new stadium. The second community (top right), consisting of Mike Brown, Marge Schott, business competition, renovations, and the new facility, which overlaps substantially with CCS's second community, generally accounts for the business community and its interests. Finally, the third community (bottom right) accounts for the public side of the building of the new stadium, including nodes such as the general public, the sports franchise game, creating the referendum, and passing the referendum, which was subject to a public vote.

In this example, the results are quite consistent with the "picture" almost literally painted by CCS (2003) on the basis of their deep knowledge of the Cincinnati controversy. The communities that they infer overlap substantially with those that our algorithm identified, even though we made no inferences from substantive knowledge, but rather allowed the eigenspectrum of the three-mode modularity matrix to determine the partition. Having illustrated our approach with this didactic example, we now turn to the results of simulations using much larger networks.

## 5. Simulation Results

The results in this section are based on thousands of four-mode network simulations. Given the number of nodes in each mode, the density of the constituent two-mode networks, and the probability of a tie occurring within a community, we simulated each two-mode network, computed the null models for each network, subtracted the null from the simulated network, aggregated the six two-mode networks into a four-mode network (of the form  $Z_B$  from above), and then determined the community structure of the four-mode network based on maximization of the modularity  $Q$  for the network. Below we report the proportion of times that modularity maximization identified the imposed community structure in the networks, but first we present more details of the simulation.

Ties in each of the two-mode networks that constitute the full four-mode network were allocated between two equally-sized communities with probability  $p$  and  $1 - p$ . In the simulations we report here, the first mode had 50 nodes (*a*), the second mode had 100 nodes (*b*), the third mode had 150 nodes (*c*), and the fourth mode had 200 nodes (*d*). Thus in the  $a \times b$  network, ties from the first half of the nodes in *a* to the second half of the nodes in *b* occurred with a probability of  $1 - p$ , as did the ties from the second half of the nodes in *a* to the first half of the nodes in *b*.

Aside from two equally-sized communities, we imposed a few other constraints for the sake of parsimony. First, the density of each of the two-mode networks that went into the full four-mode network was constrained to be equal. Second, the degree of each node in the first mode was constrained to be equal to the density times the number of nodes in the second mode. Ties were probabilistically distributed, but that does not

<sup>4</sup>An R workspace and script are available on the first-listed author's website that replicates the results reported in this paper.

imply that the degrees of each node in the second mode were equated.

We manipulated the density of the constituent two-mode networks to be between .2 and .5 in .1 intervals. We also manipulated the probability of a tie occurring within communities to be between .65 and .80 in .05 intervals. For each combination of densities and probabilities, we simulated 1,000  $Z_B$  matrices and retained the community structure associated with the maximized modularity  $Q$ . Figure 2 illustrates the proportion of times that the community structure we found matched the *a priori* partition based on the probability of within-community ties. Not surprisingly, as the probability of within-mode ties increases, so too does the proportion of times that modularity maximization identifies the ‘correct’ community structure. In sparser networks, the probability of within-community ties has a large impact on the precision of the algorithm. With network densities of .2, the probability of a within-mode tie of .65 lead to the identification of only one correct community structure, and the probability of a within-mode tie of .80 lead to the identification of 999 correct community structures.

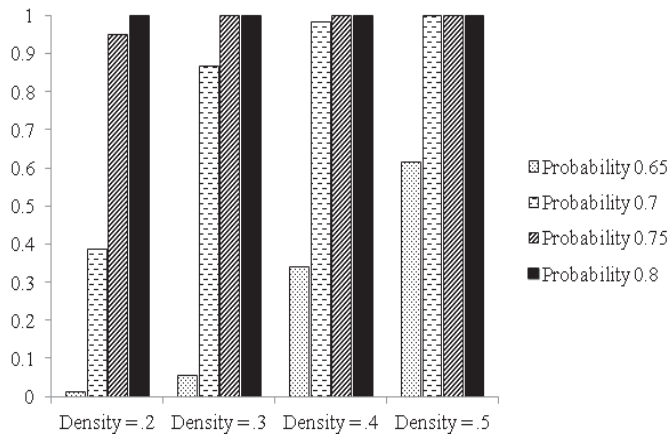


Figure 2. Simulation results presenting the proportion of times the algorithm identifies the *a priori* community partition under varying network densities and probabilities of within-community ties.

Somewhat surprisingly, network density has a reasonably strong effect on the accuracy of the algorithm. In sparser networks the probability of a within-community tie has more of an effect on the precision of the algorithm than in denser networks. For within-community ties with a probability of .7, for example, the algorithm correctly identifies 38.4% of the simulated networks’ community structure when the constituent matrices have a density of .2, but it identifies 98.2% of the simulated networks’ community structure when the constituent matrices have a density of .4. Thus, based on our simulation results, it appears that the precision of the approach outlined above is affected by the overall strength of the community structure, and the density of the networks. In retrospect, networks with more ties to probabilistically allocate within communities *should* result in more accurate community identification because there

is more information to exploit. Also, although we do not report the results of our other simulations here, they suggest that the patterns found in Figure 2 are roughly reliable for significantly larger networks.

## 6. Conclusion

We have combined the logic of multi-mode networks with modularity-based community finding using spectral partitioning of the modularity matrix. We illustrated how to construct the multi-mode network before computing its eigenstructure. We then applied this algorithm to a published example and showed the overlap of our results with CCS’s (2003) results that were based on substantial substantive knowledge. We also reported on the results of simulations of large four-mode networks, which illustrated the importance of network density for the identification of community structure.

Three points warrant further mention. First, the procedure described herein can be applied to any number of modes. We analyzed a three-mode network, simulated four-mode networks, and West et al. (2012) used this procedure with four-mode data. Second, it is possible that within-mode ties may be incorporated with between-mode ties in a manner similar to that described above. This may be accomplished, for example, by treating the one-mode network as another two-mode network in the construction of  $Z$  (i.e., maintaining the block off-diagonal form, but including within-mode ties and their transpose). Such a formulation would maintain the assumption of a symmetrical adjacency matrix, but would actually be a multi-level network. Subsequent sensitivity analyses will be required to validate whether this identifies the community structure of multi-level networks acceptably well.<sup>5</sup>

Third, there is no limit to the number of vertices that can be partitioned into communities using our approach. In our empirical example, there were only twenty nodes, enabling us to compare results to an extant substantively meaningful account (Cornwell et al., 2003). As the number of vertices in a network increases, the ability to obtain and process substantive and qualitative knowledge decreases, thus increasing the need for reliable quantitative procedures such as the one we have proposed here. In this vein, the results of our simulations show the reliability of our procedure, and suggest that network density is an important part of the community structure puzzle.

## References

- Barber, M. J. (2007). Modularity and Community Detection in Bipartite Networks. *Physical Review E*, 76, 066102.
- Brandes, U., Delling, D., Gaertler, M., Görke, R., Hoefer, M., Nikoloski, Z., & Wagner, D. (2008). On Modularity Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 20, 172-188.
- Breiger, R. L. (1974). The Duality of Persons and Groups. *Social Forces*, 53(2), 181-190.

<sup>5</sup> We thank an anonymous reviewer for bringing within-mode ties to our attention.

- Carley, K. M. (2003). Dynamic Network Analysis. In R. L. Breiger, K. M. Carley and P. E. Pattison (Eds.), *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, (pp. 133-145). Washington DC, VA: National Academies Press.
- Cornwell, B., Curry, T. J., & Schwirian, K. P. (2003). Revisiting Norton Long's Ecology of Games: A Network Approach. *City & Community*, 2, 121-142.
- Fararo, T., & Doreian, P. (1984). Tripartite Structural Analysis: Generalizing the Breiger-Wilson Formalism. *Social Networks*, 6, 141-175.
- Fielder, M. (1973). Algebraic Connectivity of Graphs. *Czechoslovak Mathematical Journal*, 23, 298-305.
- Girvan, M., & Newman, M. E. J. (2002). Community Structure in Social and Biological Networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), 7821-7826.
- Guimerà, R., Sales-Pardo, M., & Amaral, L. A. N. (2007). Module Identification in Bipartite and Directed Networks. *Physical Review E*, 76, 036102.
- Jackson, M. O. (2008). *Social and Economic Networks*. Princeton, NJ: Princeton University Press.
- Long, N. E. (1958). The Local Community as an Ecology of Games. *American Journal of Sociology*, 64, 251-261.
- Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., & Onnella, J. P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980), 876-878.
- Murata, T. (2011). A New Tripartite Modularity for Detecting Communities. *Computer Software*, 28(1), 154-161.
- Newman, M. E. J. (2006). Modularity and Community Structure in Networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23), 8577-8582.
- Palla, G., Derényi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society. *Nature*, 435, 814-818.
- Porter, M. A., Mucha, P. J., Newman, M. E. J., & Friend, A. J. (2007). Community Structure in the United States House of Representatives. *Physica A*, 386, 414-438.
- Porter, M. A., Onnella, J. P., & Mucha, P. J. (2009). Communities in Networks. *Notices of the American Mathematical Society*, 56, 1082-1097 & 1164-1166.
- Pothen, A., H. D. Simon, & K. Liou. (1990). Partitioning Sparse Matrices with Eigenvectors of Graphs. *SIAM Journal on Matrix Analysis and Applications*, 11, 430-452.
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. New York: Cambridge.
- West, A.J., Melamed, D., & Breiger, R.L. (2012). Finding Community Structure in Multi-Mode Networks: Narrative Analysis of the Afghan Mujahideen. Presented at the International Sunbelt Social Networks Conference, Redondo Beach, CA, March 16.
- Zhang, Y., Friend, A. J., Traud, A. L., Porter, M. A., Fowler, J. H., & Mucha, P. J. (2008). Community Structure in Congressional Cosponsorship Networks. *Physica A*, 387, 1705- 1712.

## Injection Drug Users' Involvement In Drug Economy: Dynamics of Sociometric and Egocentric Social Networks

---

**Cui Yang**

*Johns Hopkins  
Bloomberg School of Public Health  
Baltimore, Maryland*

**Carl Latkin**

*Johns Hopkins  
Bloomberg School of Public Health  
Baltimore, Maryland*

**Stephen Q. Muth**

*Quintus-Ential Solutions  
Colorado Springs, Colorado*

**Abby Rudolph**

*Alcohol, Policy, and Safety Research Center  
Calverton, Maryland*

### Abstract

The purpose of this analysis was to examine the effect of social network cohesiveness on drug economy involvement, and to test whether this relationship is mediated by drug support network size among a sample of. Involvement in the drug economy was defined by self-report of participation in at least one of the following activities: selling drugs, holding drugs or money for drugs, providing street security for drug sellers, cutting/packaging/cooking drugs, selling or renting drug paraphernalia (e.g., pipes, tools, rigs), and injecting drugs in others' veins. The sample consists of 273 active injection drug users in Baltimore, Maryland who reported having injected drugs in the last 6 months and were recruited through either street outreach or by their network members. Egocentric drug support networks were assessed through a social network inventory at baseline. Sociometric networks were built upon the linkages by selected matching characteristics, and k-plex rank was used to characterize the level of cohesiveness of the individual to others in the social network. Although no direct effect was observed, structural equation modeling indicated k-plex rank was indirectly associated with drug economy involvement through drug support network size. These findings suggest the effects of large-scale sociometric networks on injectors' drug economy involvement may occur through their immediate egocentric networks. Future harm reduction programs for injection drug users (IDUs) should consider providing programs coupled with economic opportunities to those drug users within a cohesive network subgroup. Moreover, individuals with a high connectivity to others in their network may be optimal individuals to train for diffusing HIV prevention messages.

### Authors

*Cui Yang* is a scientist in the department of Health, Behavior and Society at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland.

*Carl Latkin* is a professor in the department of Health, Behavior and Society at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland.

*Stephen Q. Muth* is the Director of Quintus-Ential Solutions in Colorado Springs, Colorado.

*Abby Rudolph* is a research scientist in Alcohol, Policy, and Safety Research Center in Calverton, Maryland.

### Notes

Dr. Cui Yang was funded by National Institute of Alcohol Abuse and Alcoholism (NIH/NIAAA: 1K99AA020782), Dr. Abby Rudolph was funded by National Institute of Drug Abuse (NIH/NIDA: K01DA033879). The SHIELD study was funded by the National Institute on Drug Abuse (NIDA; R01DA13142; Principal Investigator: Carl Latkin).

*Please address correspondence to Cui Yang, Johns Hopkins Bloomberg School of Public Health, 2213 McElderry Street, 2nd Floor, Baltimore, MD 21205, USA. Phone: 410-502-5368/ Fax: 410-502-5385/ E-mail: cyang@jhsph.edu*

## 1. Introduction

### 1.1 Background

Drug economy, defined as a range of drug-related behaviors an individual engages in for financial purpose (e.g. selling drugs or drug paraphernalia), offers a source of income in the absence of licit employment opportunities for many marginalized populations, including illicit drug users (Dunlap et al., 2010; Gwadz et al., 2009; Latkin, Davey, & Hua, 2006). Participation in the informal economy was demonstrated by Gwadz and colleagues (2009) to be related to the presence of strong ties to others involved in the informal economy and a perception of the informal economy as a norm (Gwadz et al., 2009). Further, among drug users, information on buying drugs is often obtained through social networks, and drugs are often purchased and used with other users (Latkin et al., 2006). Likewise, drug use among network members may influence one's participation in the drug economy. For example, a study among Thai methamphetamine drug users reported that having a greater proportion of drug network members who recently stopped using methamphetamine was associated with decreased likelihood of participating in the drug economy (Latimore et al., 2011). These studies collectively suggest the important role that peers play in influencing one's participation in the drug economy. Because both individual drug use and peer drug use increase one's connections to those already involved in the drug economy, one's likelihood of participation in the drug economy is dependent on individual, network, and structural factors. In this analysis, we aim to assess the composition and structure of drug user networks and to identify network features which are indirectly and directly associated with drug economy involvement among a sample of active injection drug users in Baltimore, Maryland.

### 1.2 Drug economy

Patterns of drug use are influenced by macro and micro economic factors. Since the 1970s, increasing deindustrialization has brought socioeconomic deprivation in urban settings in the United States, characterized by plant closings, massive job losses, and population instability (Bluestone & Harrison, 1988). Urban poverty has been associated with higher rates of illicit drug use and HIV infection, which have disproportionately affected minority groups, such as African Americans (McCord & Freeman, 1990; Wallace, 1990). The costs associated with illicit drugs force drug users, particularly those with severe drug addiction, to engage in drug economy activities (Debeck et al., 2007), and the availability of drugs among individuals in the drug economy may foster drug use.

Involvement in the drug economy places drug users at risk of violence and incarceration (Curry & Latkin, 2003b; Sherman & Latkin, 2002b; ). In a sample of street-involved youths, involvement in the drug trade was associated with homelessness and self-reported police assault (Werb, Kerr, Li, Montaner, & Wood, 2008). A gender difference has been observed with respect to the types of drug economy activities for which men and women are arrested. Male heroin injectors were more likely to get arrested for selling drugs, while steering/touting

(publicizing) drugs was associated with female injectors' arrest (Curry & Latkin, 2003a). Involvement in the drug economy can also increase health-related risk behaviors. Friedman and colleagues found that IDUs involved in the drug economy were more likely to have HIV and other blood-borne infections as compared to those drug injectors not involved in the drug economy (Friedman et al., 1998). Network characteristics, structure, ties to the drug economy, and social norms about involvement in the drug economy may also influence whether an individual participates in the drug economy. Therefore, it is important to understand the network characteristics, peer connections, and network structure which make drug users more likely to participate in and to inform public health strategies to improve the lives and well-being of drug users.

### 1.3 Social network analysis (SNA)

Two types of networks are discussed in the literature: risk networks and social networks. Risk networks consist of those engaged in risk behaviors and social networks are comprised of individuals providing social support. There are two fundamental analytic approaches in SNA: egocentric and sociometric. The egocentric approach focuses on respondents' direct personal ties, usually relying entirely on respondents' self report of behaviors and attributes for those ties; the sociometric approach describes a larger set of relationships — the entire panoply of linkages among multiple respondents (Wasserman & Faust, 1994). A fundamental difference between egocentric and sociometric analysis is that each respondent's immediate group is considered independent in egocentric network data analysis, whereas the entire network is considered to be the unit of analysis with sociometric data (Wasserman & Faust, 1994).

Egocentric network characteristics, such as network size and composition, have been linked to a number of drug-related behaviors, including sharing injection equipment (Costenbader, Astone, & Latkin, 2006; Lakon, Ennett, & Norton, 2006; Latkin et al., 1995; Suh, Mandell, Latkin, & Kim, 1997), exchanging sex for money or drugs (Latkin, Hua, & Forman, 2003), overdose (Tobin, Hua, Costenbader, & Latkin, 2007), and entry to drug treatment (Davey, Latkin, Hua, Tobin, & Strathdee, 2007). A few studies have examined the association between egocentric network characteristics and drug economy involvement. In a sample of active IDUs in Baltimore, Sherman and colleagues found that drug users involved in the drug economy were likely to have more daily contact with drug users, as well as to have a greater percentage of drug users in their social networks (Sherman & Latkin, 2002a). Similarly a study among Thai methamphetamine users reported a positive association between the total number of methamphetamine using networks and one's involvement in the drug economy and an inverse association between the proportion of methamphetamine using networks who recently quit using methamphetamine and one's involvement in the drug economy (Latimore et al., 2011).

Network structure may also influence one's likelihood of participating in the drug economy. Behaviors, information, and disease may flow more easily through dense (or more cohesive) networks because there are more paths connecting any two members of a network. (Liljeros, Edling et al., 2003; Wasser-



man & Faust, 1994). For example, a greater number of connections to drug users may increase one's likelihood of participating in the drug economy by virtue of the increased number of paths for entry. Networks can also influence drug-related and sex-related risk behaviors through social norms (Davey-Rothwell & Latkin, 2007; De, Cox, Boivin, Platt, & Jolly, 2007; Latkin et al., 2004; Latkin et al., 2006; Shaw et al., 2007; Tobin et al., 2010; Unger et al., 2006). The perception of involvement in the drug economy as a norm (in networks where drug economy involvement is high) may also increase one's participation in the drug economy.

Examining group structure provides additional insights over those provided merely by assessing attributes of individuals. To this end, various classes of sociometric network metrics have been developed either to characterize network systems as a whole, or to characterize the significance of the individual node in a network context (Wasserman & Faust, 1994). Network measures of centrality such as betweenness (Freeman, 1979), information centrality (Stephenson & Zelen, 1989), and eigenvector centrality (Bonacich, 1987) are well-suited to such tasks. Previous studies documented that sociometric characteristics are associated with drug equipment sharing, and HIV/STI transmission (Curtis et al., 1995; Friedman et al., 1997; Rothenberg, Hoang, Muth, & Crosby, 2007). For example, Friedman and colleagues found an association between position as core members of an IDU network and drug equipment sharing behaviors and HIV acquisition (Friedman et al., 1997).

One of the limitations inherent in sociometric network analysis is that many network metrics (e.g. most distance-based centrality measures) apply only within connected "components" – groups where all nodes are reachable by a path of some length. When high connectivity is not assured because of difficulties in identification of alters or a sampling methodology that does not place a premium on recruiting linked persons, it is still possible to use metrics of subgroup formation, such as cliques (Luce & Perry, 1949) and k-plexes (Seidman & Foster, 1978) to rank network members according to the structural complexity of "microstructures" in a network. We used k-plex ranks in the present analysis for its ability to capture the extent of each person's participation in microstructures, where each member is connected to at least n-k other members within the group (Seidman et al., 1978). As the size of a k-plex increases, so does the sheer number of connections required to maintain the k-plex; in other words, the network cohesion also increases. Since the extent of involvement in these cohesive subgroups is not dependent on which connected component the egos belong to, k-plex rank provides a more robust measure of their network connectivity relative to other sociometric measures (e.g., betweenness, information centrality, or eigenvector centrality).

Despite the differences between egocentric and sociometric network analyses, they are not mutually exclusive. Egocentric network data have been used to build sociometric networks to assess a variety of health-related topics, including smoking cessation, obesity, drug injection behaviors and HIV/STI transmission (Christakis & Fowler, 2007; Christakis & Fowler, 2008; Friedman et al., 1997; Rothenberg et al., 1998). Sociometric network measures are able to assess fundamental social structures that cannot be reduced to individual level factors. It

is likely that many behaviors are influenced by direct network contacts, such as family members, friends, or drug-sharing partners, and those can provide proximate potential resources. However, previous studies have rarely utilized both methods of analyses (Gyarmathy & Neaigus, 2006).

A better understanding of how egocentric and sociometric networks interplay to affect risk behaviors may allow for developing more effective interventions (De et al., 2007). The goal of the current analysis is to model the potential causal pathway between sociometric network features, egocentric, network characteristics, and participation in the drug economy. In this paper, we hypothesized that sociometric network features such as k-plex rank, have direct and indirect effects, mediated through constituent egocentric network characteristics (number providing drug support), to involvement in the drug economy among a sample of active IDUs in Baltimore, Maryland. This hypothesis was tested with structural equation models (SEM) using the latent variable of drug economy involvement. The approach used in this analysis attempts to tie egocentric networks together to assess the larger social network they create, to better understand the social structure of drug economy involvement.

## 2. Methods

### 2.1 Data source

The survey data used in this analysis were collected as a part of the SHIELD (Self-Help In Eliminating Life-threatening Diseases) project, a network-oriented experimental pre- and post-test intervention. Index participants were recruited through targeted outreach in high drug use areas. SHIELD study inclusion criteria consisted of: 1) being at least 18 years old, 2) having daily or weekly contact with drug users, 3) willingness to conduct AIDS outreach education, 4) being able to bring in 2 network members for a baseline interview, and 5) not being enrolled in other HIV prevention or network studies. Index participants were asked to bring 2 high-risk members of their networks to the clinic for assessment after the initial interview. Network members were eligible if they were at least 18 years old and were referred by an eligible index participant. Face-to-face and ACASI interviews were conducted to assess their sociodemographic characteristics, HIV related behaviors, and their social networks. There were five waves of data collection from 1997 through 2004. The Johns Hopkins School of Public Health Committee on Human Research approved the study.

## 3. Measures

### 3.1 Network configuration

The egocentric network characteristics were assessed at three time points (waves 1, 2, and 4), using the Personal Network Inventory, a modified version of the Arizona Social Support Inventory (Barrera, 1981). This inventory has been shown to have good concurrent and predictive validity and internal consistency (Latkin et al., 1996). The first section of this inventory had nineteen questions designed to generate names of people

in respondents' personal networks, including persons providing or receiving social support, using drugs (injecting or not), or sex partners in the past six months. Characteristics of each nominated network member, such as age, relationship, frequency of contact, duration of relationship and types of drug use were also assessed. The primary egocentric network characteristic in this analysis was the size of each respondent's drug support network at the baseline, assessed by four name generating questions including: "Who you do drugs with?", "Who do you consider your walking partner/running buddy?", "If you were going through withdrawal, who can/could you usually count on to get you drugs?", and "Who can/could you usually count on for drugs or the money to get drugs?"

Sociometric networks were built upon the existing information of respondents (egos) and nominated individuals (alters) who have been interviewed in all three waves (waves 1, 2 and 4,  $N=611$ ), 315 (51.6%) of which were index participants and 296 (48.5%) were recruited network members. The sociometric linkages were confirmed by selected matching characteristics, such as first and last name, age, gender, and address. Alters' full names were only available in wave 4 after the Committee on Human Research approval. As a consequence, matching alters' identities was better facilitated in wave 4, resulting in a bias against ascertainment of larger network structure for respondents interviewed exclusively in earlier waves. Additionally, respondents interviewed less than three times had fewer chances to provide alters, therefore we restricted analysis to participants interviewed in all three waves, providing a more consistent and less biased sociometric network sample.

Four different levels of matching certainty were applied as "certain," "probable," "possible" and "improbable." Moreover, four broad categories of name generators, equipment, drug, sexual and social were used to elicit identities of alters. Equipment alters were persons with whom the ego shared, borrowed or lent either needles or cookers. Drug alters were persons with whom the ego did drugs. Sex alters were persons with whom the ego had sex in the last 6 months, and social alters included those elicited by questions, such as "Give me the first name, and last name initial of people who you would talk to about things that are very personal and private?" "Is there anybody that you could get together with to have fun or to relax or just hang out with?" In preliminary analyses, twelve networks were examined based on four combinations of linkage attributes (equipment-only, sex-equipment, sex-drug-equipment and all), each under three match certainty assumptions (certain, probable and possible). In the current analysis, we used the network composed of all types of alters under the most conservative (i.e. "certain") assumption of matching certainty.

K-plex ranks were calculated using UCINET (Borgatti, Everett, & Freeman, 2002) and SAS (SAS Institute, 2011). We enumerated 2-plexes of all sizes with UCINET, and post-processed this information with SAS to create the ranked k-plex score with a scheme similar to one used in a network study of linkages among people and places in a TB investigation (Cook et al., 2007). The network visualization software Pajek was used to create network images (Batagelj & Mrvar, 1998).

### 3.2 Drug economy

Drug economy involvement was assessed at the baseline. Respondents were asked if they had performed at least one of the following seven roles in the six months prior to the baseline interview: 1) sold drugs; 2) steered customers to or touted (publicized) drugs; 3) held drugs or money for drugs; 4) provided street security for drug sellers including being a "lookout" for police; 5) cut, packaged, or cooked drugs; 6) sold or rented pipes/tools/rigs and 7) "street doctored" (inject into the veins of others).

### 3.4 Sociodemographic and drug use characteristics

Baseline sociodemographic characteristics examined in this analysis were race/ethnicity (African-American vs. others), gender, age, education (at least high school diploma or GED), relationship status (currently having main partner vs. others), current employment, monthly income (median split for \$1,000 or more), source of income, homelessness, and history of arrest in the past year. Respondents reported on the frequency of injecting heroin, cocaine and speedball (i.e., a combination of heroin and cocaine) in the past 6 months, and daily injectors were operationalized as respondents who have injected heroin, cocaine or speedball at least every day in the past 6 months.

## 4. Data analysis

The current analysis was limited to SHIELD participants who had injected heroin, cocaine or speedball within the six months prior to the baseline data collection, and been regularly-interviewed in waves 1, 2, and 4, from 1997 through 2003 ( $N=273$ ).

The construct validity of the drug economy scale was evaluated. First, an exploratory factor analysis (EFA) of the correlation matrix of the original seven items was analyzed with the Mplus program 5.21 (Muthen & Muthen, 2007). As variables for the drug economy were categorical, the mean and variance-adjusted weighted least-squares estimator was used. The factors were correlated under the oblique geomin rotation. Two criteria were used to determine the number of factors to be extracted in the exploratory factor analysis model: 1) the number of eigenvalues greater than one and 2) the scree plot (Netemeyer, Bearden, & Sharma, 2003). Sizes of the loading and cross loadings were examined to determine the quality of the variables measuring the factors. A confirmatory factor analysis (CFA) was then conducted to examine the fit of the factor solution using the items chosen from EFA. Goodness-of-fit was evaluated by five indices: the standardized root mean residual (SRMR) is close to .08 or below, the weighted root-mean-square residual (WRMR), is 1.00 or below, the root-mean-square-error approximation (RMSEA) is close to .06 or below, and the comparative fit index (CFI), and the Tucker-Lewis Index (TLI) values close to .95 or higher (Hu & Bentler, 1999).

A composite score for the drug economy activities was generated by adding dichotomized responses from selected items from EFA and CFA. A binary variable for drug economy involvement was created for the descriptive analysis. Bivariate

analyses were conducted to compare characteristics of active IDUs involved in any drug economy activities at the baseline to those not involved at baseline. Tests for significance of differences in proportions were used for categorical variables. For continuous variables, analysis of variance was used for normally distributed variables, and Kruskal–Wallis tests for non-normally distributed variables. Data were analyzed using Stata 10.0 (StataCorp., 2005).

To test the hypothesis that sociometric network characteristics (k-plex rank) are associated with the drug economy involvement directly and indirectly through egocentric network characteristics (number of network members providing drug support), we conducted structural equation modeling (SEM) techniques using Mplus. We tested a model in which k-plex rank had a direct effect on both the number of drug support network members and drug economy involvement, and a model in which k-plex rank had both a direct and an indirect effect on drug economy involvement through the number of drug support network members. Other independent variables (i.e., gender and daily injectors) previously found to be associated with drug economy involvement were included in the model. Given the small sample size and non-normal distributions of the mediator and outcome, the bootstrap option was used to estimate the standard errors (Shrout & Bolger, 2002). Model fit was evaluated with RMSEA and WRMR.

## 5. Results

### 5.1 Drug economy

Among 1,637 participants in the SHIELD baseline, 273 injectors were interviewed at waves 1, 2, and 4. In the EFA of the drug economy scale, two factors with eigenvalues of over 1.00 were identified, and examination of the scree plot confirmed a two-factor solution. Table 1 presents sizes of the loading of the items measuring the factors and model fit indices (SRMR=0.045, RMSEA=0.053, CFI=0.997, TLI=0.991). Item 2 (“steered customers to or touted [publicized] drugs”) had

cross-loading on both factors. In accordance with the recommendation that items with cross-loading less than 0.15 difference from its highest factor should be deleted (Worthington & Whittaker, 2006), item 2 was removed from the scale. A two-factor model with latent constructs representing drug-selling activities (sold drugs; held drugs or money for drugs; provided street security for drug sellers including being a “lookout” for police; cut, packaged, or cooked drugs) and injection-related activities (sold or rented pipes/tools/rigs; street doctor or injecting the veins of others) was used for further analyses. Fit indices of CFA indicated good model fit for a two-factor solution of 6 selected items from the EFA (WRMR=0.558, RMSEA=0.043, CFI=0.996, TLI=0.993).

The median number of drug economy activities was 1 (mean: 1.38, range 0 – 6). More than half of the sample (54.6%) had at least one drug economy activity in the past 6 months. Table 2 compares the characteristics of IDUs with at least one drug economy activity with those without drug economy activity. Being involved in drug economy activity was associated with hustling and having a friend/family/sexual partners as sources of income, and injecting drugs daily. In addition, larger drug support networks were associated with drug economy involvement.

Figure 1 presents the visualization of the entire network of sex, equipment, drug, and social connections (N=273 respondents), while Figure 2 provides a close-up of the eleven components that contain higher-order microstructure (N=62 respondents). Respondents are visualized as open squares (males) and circles (females), proportional to their reported number of drug economy activities (see Legend). Non-respondent alters are the smallest (closed) squares and circles depicted in the figures. The respondents depicted in Figure 2 (the upper left quadrant of Figure 1) have higher k-plex rank scores than the balance of the respondents, due to greater involvement in network microstructures of higher complexity. Generally, respondents with higher k-plex rank are positioned closer to the upper left in both figures. The three respondents with highest k-plex rank are indicated with arrows in the Figure 2.

**Table 1.** Exploratory factor analysis (EFA) of drug economy sale.

Items	Factor 1	Factor 2
1) Sold drugs	0.902	0.000
2) Steered customers to or touted (publicized) drugs	0.513	0.511
3) Held drugs or money for drugs	0.848	0.161
4) Provided street security for drug sellers which includes being a “lookout” for police	0.636	0.350
5) Cut, packaged, or cooked drugs	0.951	-0.231
6) Sold or rented pipes/tools/rigs	-0.008	0.726
7) Street doctor or hitting veins on others	0.058	0.667
<b>Model fit indices</b>		
SRMR		0.045
RMSEA		0.053
CFI		0.997
TLI		0.991

**Table 2.** Sociodemographics, drug use behaviors, personal network and sociometric network characteristics of active IDUs, SHIELD Study (N=273).

	Total (n=273) N (%)	No drug economy role (n=124) N (%)	At least one drug economy role (n=149) N (%)	p
Age				
≤42	159 (58)	69 (56)	90 (60)	0.43
>42	114 (42)	55 (44)	59 (40)	
Race				
Other	14 (5)	4 (3)	10 (7)	0.19
African American	259 (95)	120 (97)	139 (93)	
Gender				
Male	175 (64)	76 (61)	99 (66)	0.38
Female	98 (36)	48 (39)	50 (34)	
Education				
Less than high school	139 (51)	66 (53)	73 (49)	0.49
At least high school graduate	134 (49)	58 (47)	76 (51)	
Currently having main partner				
No	98 (36)	40 (32)	58 (39)	0.25
Yes	175 (64)	84 (68)	91 (61)	
Current employment				
Unemployed	236 (86)	104 (84)	132 (89)	0.26
Employed	37 (14)	20 (16)	17 (11)	
Monthly income				
≥\$1,000	20 (7)	14 (9)	6 (5)	0.15
<\$1,000	253 (93)	135 (91)	118 (95)	
Source of income				
Salaries or wages	47 (17)	24 (19)	23 (15)	0.39
Welfare/public assistance	93 (34)	38 (31)	55 (37)	0.28
Food stamps	141 (52)	63 (51)	78 (52)	0.80
Social security	57 (21)	30 (24)	27 (18)	0.22
Hustling(legal or illegal)	119 (44)	41 (33)	78 (52)	0.001
Friends/family/sexual partner	159 (58)	61 (49)	98 (66)	0.006
Homeless <sup>#</sup>				
No	230 (85)	105 (85)	125 (84)	0.96
Yes	42 (15)	19 (15)	23 (16)	
Incarcerated*				
No	226 (83)	107 (86)	119 (80)	0.16
Yes	47 (17)	17 (14)	30 (20)	
Daily injectors <sup>#</sup>				
No	104 (38)	57 (46)	47 (32)	0.02
Yes	170 (62)	68 (54)	102 (68)	
Size of drug support network: Mean (SD)	4.9 (2.8)	4.2 (2.3)	5.4 (3.1)	<.001
k-plex rank: Mean (SD)	156.5 (102.1)	154.1 (108.8)	158.5 (96.4)	0.72

### 5.2 Structural equation model

The model specified a pathway from k-plex rank through the drug support network and to two measures of drug economy, fitting the data well (RMSEA=0.000, WRMR=0.574). Standardized path coefficients were presented to facilitate comparisons among the coefficients. Figure 3 presents the complete model with standardized path coefficients. The factor correlation for drug-selling and injection-related activities was 0.58 ( $p<.001$ ). There was no significant direct effect from k-plex rank to either of the two measures of drug economy. K-plex rank had significant indirect effects on both drug selling ( $\beta=0.076$ ) and injection-related activities ( $\beta=0.081$ ). Females were less likely to be involved in drug selling activities, while daily injectors were more likely to get involved in both drug selling and injection-related activities. For the drug support network, the R2 (variance explained) was 0.18; for drug-selling

activities, the R2 was 0.13, and for injection-related activities, the R2 was 0.14. The final model provided support for the hypothesis that k-plex rank was indirectly associated with the drug economy activities through the presence of more drug support networks.

## 6. Discussion

Using structural equation modeling, we examined the relationships among sociometric network characteristics, egocentric network composition and the drug economy involvement. Although no direct effects were observed, k-plex rank was indirectly associated with two measures of drug economy involvement through the size of their drug support network. IDUs in highly connected social-drug-sex-equipment networks may have frequent direct access to drug support networks, leading to the increased likelihood of being involved in the drug econ-

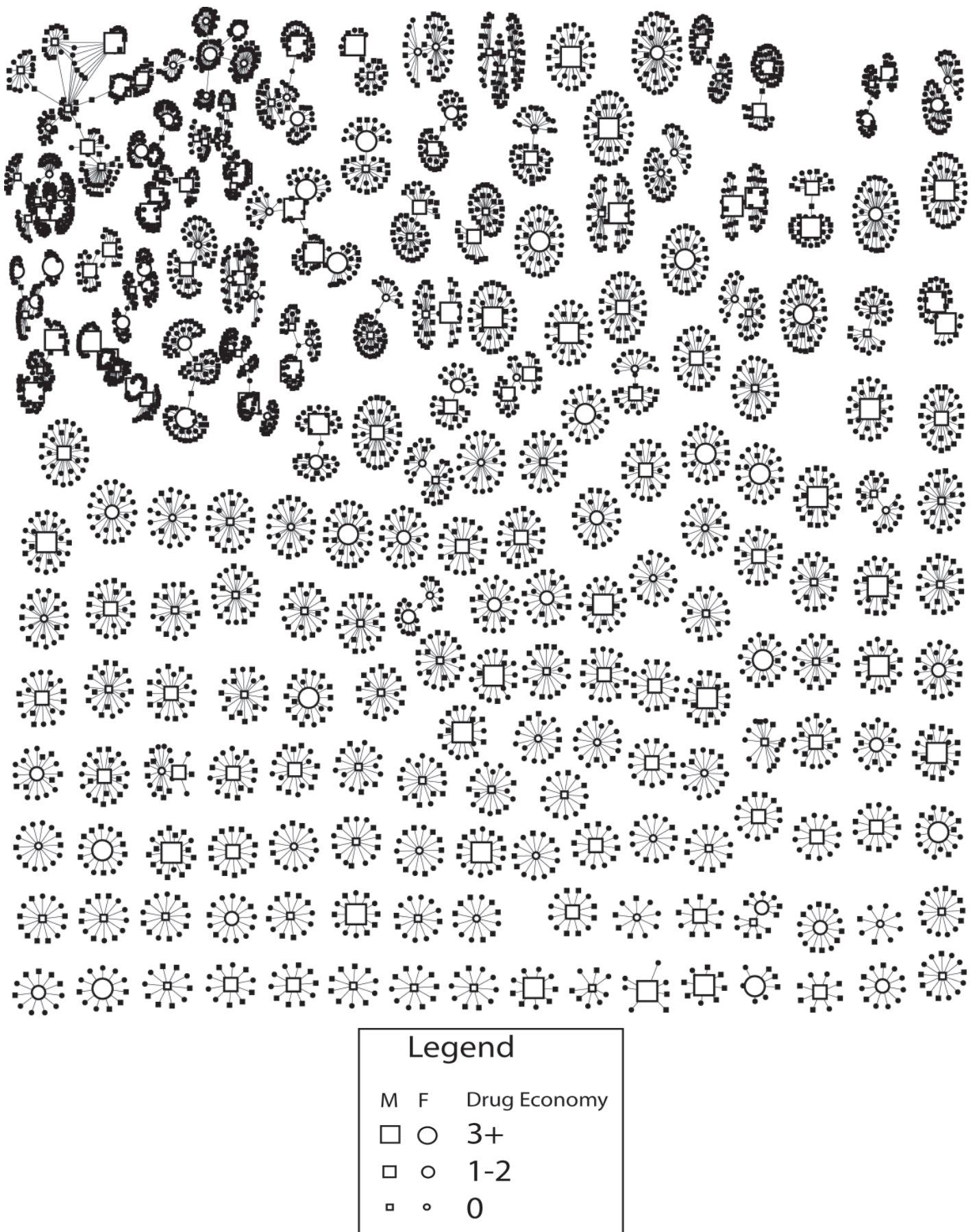


Figure 1. SHIELD study network of recent heroin and crack injectors, Baltimore Maryland, USA (N=273).

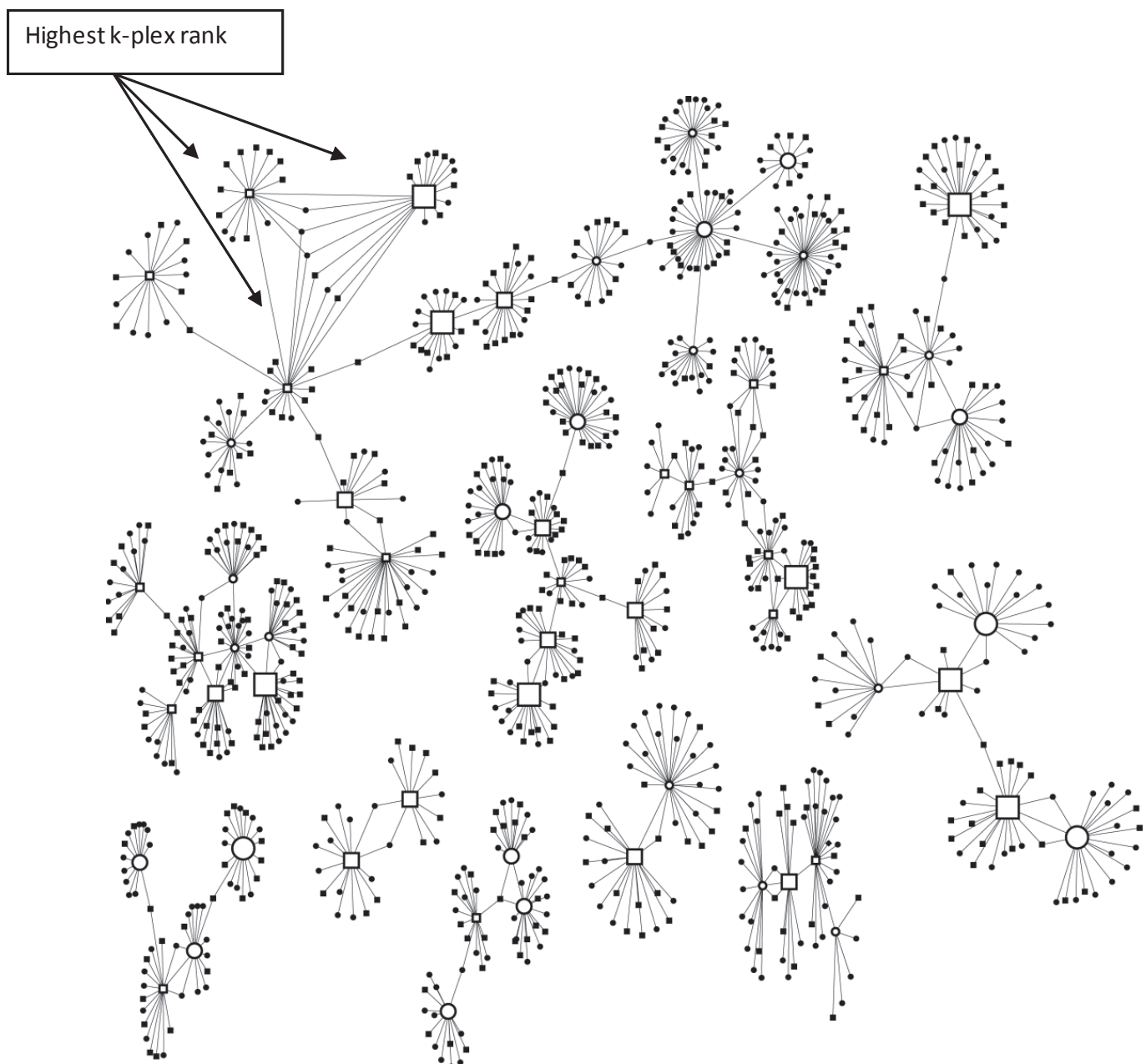


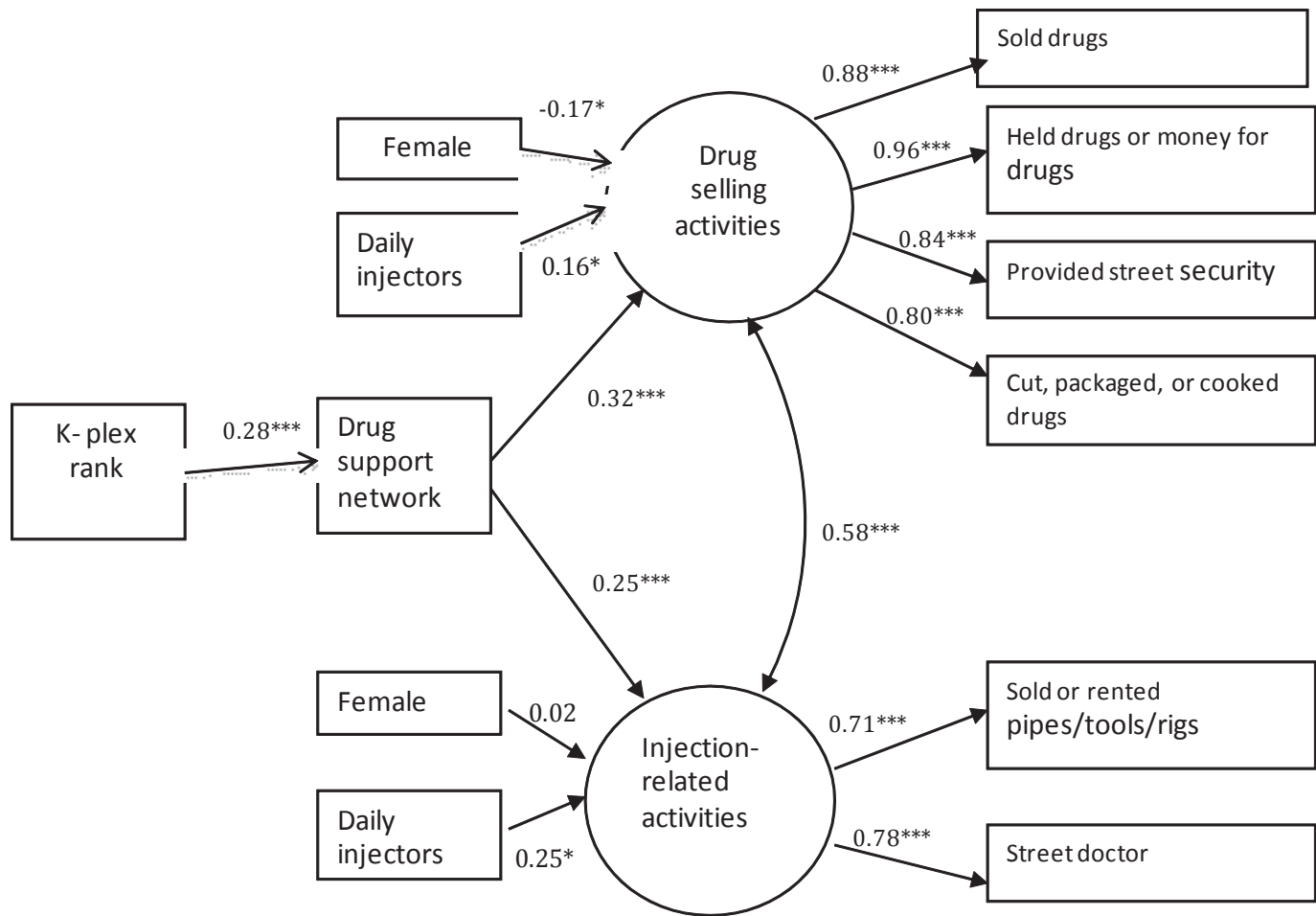
Figure 2. Eleven components of the SHIELD study network, showing respondents with higher k-plex rank (N=62).

omy to sustain their drug addiction. The current findings may also suggest that one's most immediate networks – i.e. those named in a network inventory, are likely have a larger influence on one's participation in the drug economy than those who are more loosely connected to him/her.

This analysis has several limitations. Due to the study design, the sequence of causal pathway cannot be established from k-plex rank to drug economy involvement through the drug support network. An alternative explanation is that through the drug economy, IDUs frequently interact with other drug users, which may lead to increasing cohesiveness of the risk networks. Additionally, generalizability of the findings is restricted due to the sampling strategy and analytical strategy. Identity-matching of alters was based on the completeness of various combinations of selected matching characteristics, with

full name of alters unavailable until wave 4. Any incompleteness of information undoubtedly led to missed matches and consequent underassessment of network complexity. Sociometric networks were built upon the existing information of respondents (egos) and nominated individuals (alters) who have been interviewed in all three waves. Those participants who lost to follow-up were more likely to be arrested or involved in the drug economy. In addition, the face-to-face assessment of high risk behaviors, such as drug economy involvement, may have the potential for heightened social desirability response bias. Finally, sociometric networks were comprised of networks listed at three different time points, whereas drug support networks were listed only at baseline. However, we expect the social network was relatively stable.

The present study findings provide a better understanding



**Notes**

Root mean Square Error of Approximation (RMSEA) Estimate 0.000  
 Weighted Root Mean Square Residual (WRMR) Value 0.569

Indirect effects of k-plex rank

K-plex rank → drug support network → drug selling activities=0.076\*\*

K-plex rank → drug support network → injection-related activities=0.081\*\*

Direct effects of k-plex rank

K-plex rank → drug selling activities=0.077 (NS)

K-plex rank → injection-related activities=-0.016 (NS)

NS p>.05 (z test), \* p<.05 (z test), \*\*p<.01 (z test), \*\*\*p<.001(z test)

Figure 3. Structural equation model.

of the interplay between network structure, network composition and drug-related outcomes. Although egocentric networks are easier and less costly to investigate, they describe network characteristics only from the perspective of the ego in isolation. Despite the disconnected nature of the network sample, the current study demonstrates the utility of constructing sociometric networks, providing additional context from interactions among groups of persons – an improvement over the traditional assessment of social network. The data support the hypothesis that the effects of sociometric networks on individuals may occur through their immediate egocentric networks (Friedman & Aral, 2001). Moreover, without demonstrable linkages between all network components, it is imperative to assess appropriate network metrics that apply across components, such as those involving microstructural complexity.

Both network structure and composition can be used to assess the adoption of risk-reduction messages, norms, and social support in a cohesive network, identifying subgroups at potentially higher risk, locating targets for prevention and disrupting chains of disease transmission. Future harm-reduction interventions could be targeted in a network-informed manner, for instance, prioritizing programs with job creation and training to those drug users holding positive roles within their network subgroups. Moreover, such individuals who happen to be adjacent to dense risk microstructures might be given priority for training in disseminating risk-reduction messages and mobilizing normative pressures against high-risk behaviors.

## References

- Barrera, M. A. (1981). Social support in the adjustment of pregnant adolescents: assessment issues. In B.H.Gottlieb (Ed.), *Social Networks and Social Support* (pp. 69-96). Beverly Hills, CA: Sage.
- Batagelj, V. & Mrvar, A. (1998). Pajek-program for large network analysis. *Connections*, 21(2), 47-57.
- Bluestone, B., & Harrison, B. (1988). *The Great U-Turn: Corporate Reconstructing and the Polarizing of America*. New York, NY: Basic Books.
- Bonacich, P. (1987). Power and centrality: a family of measures. *American Journal of Sociology*, 92, 1170-1182.
- Borgatti, S., Everett, M., & Freeman, L. (2002). *UCINET 6 for Windows* [Computer software and Manual]. Harvard, MA: Analytic Technologies.
- Christakis, N. A. & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357, 370-379.
- Christakis, N. A. & Fowler, J. H. (2008). The collective dynamics of smoking in a large social network. *New England Journal of Medicine*, 358, 2249-2258.
- Cook, V. J., Sun, S. J., Tapia, J., Muth, S. Q., Argueello, D. F., Lewis, B. L., Rothenberg, R.B., & McElroy, P.D. (2007). Transmission network analysis in tuberculosis contact investigations. *American Journal of Infectious Diseases*, 196, 1517-1527.
- Costenbader, E. C., Astone, N. M., & Latkin, C. A. (2006). The dynamics of injection drug users' personal networks and HIV risk behaviors. *Addiction*, 101, 1003-1013.
- Curry, A. D. & Latkin, C. A. (2003a). Gender differences in street economy and social network correlates of arrest among heroin injectors in Baltimore, Maryland. *Journal of Urban Health-Bulletin of the New York Academy of Medicine*, 80, 482-493.
- Curry, A. D. & Latkin, C. A. (2003b). Gender differences in street economy and social network correlates of arrest among heroin injectors in Baltimore, Maryland. *Journal of Urban Health-Bulletin of the New York Academy of Medicine*, 80, 482-493.
- Curtis, R., Friedman, S. R., Neaigus, A., Jose, B., Goldstein, M., & Ildefonso, G. (1995). Street-Level Drug Markets - Network Structure and HIV Risk. *Social Networks*, 17, 229-249.
- Davey, M. A., Latkin, C. A., Hua, W., Tobin, K. E., & Strathdee, S. (2007). Individual and social network factors that predict entry to drug treatment. *American Journal on Addictions*, 16, 38-45.
- Davey-Rothwell, M. A., & Latkin, C. A. (2007). Gender differences in social network influence among injection drug users: Perceived norms and needle sharing. *Journal of Urban Health-Bulletin of the New York Academy of Medicine*, 84(5), 691-703.
- De, P., Cox, J., Boivin, J. F., Platt, R. W., & Jolly, A. M. (2007). The importance of social networks in their association to drug equipment sharing among injection drug users: a review. *Addiction*, 102, 1730-1739.
- Debeck, K., Shannon, K., Wood, E., Li, K., Montaner, J., & Kerr, T. (2007). Income generating activities of people who inject drugs. *Drug and Alcohol Dependence*, 91, 50-56.
- Dunlap, E., Johnson, B. D., et al. (2010). Macro-level social forces and micro-level consequences: Poverty, alternate occupations, and drug dealing. *Journal of Ethnicity in Substance Abuse*, 9(2), 115-127.
- Freeman, L. (1979). Centrality in social networks: conceptual clarification. *Social Networks*, 1, 215-239.
- Friedman, S. R. & Aral, S. (2001). Social networks, risk-potential networks, health, and disease. *Journal of Urban Health-Bulletin of the New York Academy of Medicine*, 78, 411-418.
- Friedman, S. R., Furst, R. T., Jose, B., Curtis, R., Neaigus, A., Des, J., Goldstein, M.F., & Ildefonso, G. (1998). Drug scene roles and HIV risk. *Addiction*, 93, 1403-1416.
- Friedman, S. R., Neaigus, A., Jose, B., Curtis, R., Goldstein, M., Ildefonso, G., Rothenberg, R., & Des Jarlais D. (1997). Sociometric risk networks and risk for HIV infection. *American Journal of Public Health*, 87, 1289-1296.
- Gyarmathy, V. & Neaigus, A. (2006). The effect of personal network exposure on injecting equipment sharing among IDUs in Budapest, Hungary. *Connections*, 27, 25-38.
- Gwadz, M. V., Gostnell, K., Smolenski, C., Willis, B., Nish, D., Nolan, T. C., Tharaken, M., & Richie, A.S. (2009). The initiation of homeless youth into the street economy. *Journal of Adolescence*, 32(2), 357-377.
- Hu, L. & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structural analysis: conventional criteria versus new alternatives. *Structure Equation Modeling*, 6, 1-55.
- Lakon, C. M., Ennett, S. T., & Norton, E. C. (2006). Mechanisms through which drug, sex partner, and friendship network characteristics relate to risky needle use among high risk youth and young adults. *Social Science and Medicine*, 63, 2489-2499.
- Latimore, A. D., Rudolph, A., German, D., Sherman, S.G., Srijohn, B., Aramrattana, A., & Celentano, D. D. (2011). Predictors of incident and recurrent participation in the sale or delivery of drugs for profit amongst young methamphetamine users in Chiang Mai Province, Thailand, 2005-2006. *International Journal of Drug Policy*, 22(4), 259-266.
- Latkin, C., Mandell, W., Oziemkowska, M., Celentano, D., Vlahov, D., Ensminger, M., & Knowlton, A. (1995). Using social network analysis to study patterns of drug use among urban drug users at high risk for HIV/AIDS. *Drug and Alcohol Dependence*, 38, 1-9.
- Latkin, C., Mandell, W., Vlahov, D., Oziemkowska, M., & Celentano, D. (1996). People and places: behavioral settings and personal network characteristics as correlates of needle sharing. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 13, 273-280.
- Latkin, C. A., Davey, M. A., & Hua, W. (2006). Social context of needle selling in Baltimore, Maryland. *Substance Use Misuse*, 41, 901-913.



- Latkin, C. A., Hua, W., & Forman, V. L. (2003). The relationship between social network characteristics and exchanging sex for drugs or money among drug users in Baltimore, MD, USA. *International Journal of STD and AIDS*, *14*, 770-775.
- Latkin, C. A., M. A. Davey, et al. (2006). Social context of needle selling in Baltimore, Maryland. *Substance Use & Misuse* *41*(6-7), 901-913.
- Latkin, C. A., Hua, W., et al. (2004). Social network correlates of self-reported non-fatal overdose. *Drug and Alcohol Dependence*, *73*(1), 61-67.
- Liljeros, F., Edling, C. R., Mimes Amaral, L.A.. (2003). Sexual networks: implications for the transmission of sexually transmitted infections. *Microbes and Infection* *5*(2), 189-196.
- Luce, R. & Perry, A. (1949). A method of matrix analysis of group structure. *Psychometrika*, *14*, 95-116.
- McCord, C. & Freeman, H. P. (1990). Excess mortality in Harlem. *New England Journal of Medicine*, *322*, 173-177.
- Muthén L. K., Muthén B. O. (2007) MPlus (Version 5.0) [Computer software]. Los Angeles, CA.
- Netemeyer, R., Bearden, W., & Sharma, S. (2003). Scaling procedures: Issues and applications. Thousand Oaks, California: Sage Publications, Inc..
- Rothenberg, R., Hoang, T. D. M., Muth, S. Q., & Crosby, R. (2007). The Atlanta urban adolescent network study: A network view of STD prevalence. *Sexual Transmitted Diseases*, *34*, 525-531.
- Rothenberg, R. B., Potterat, J. J., Woodhouse, D. E., Muth, S. Q., Darrow, W. W., & Klovdahl, A. S. (1998). Social network dynamics and HIV transmission. *Aids*, *12*, 1529-1536.
- Seidman, S. & Foster, B. (1978). A graph-theoretic generalization of the clique concept. *Journal of Mathematical Sociology*, *6*, 154.
- Shaw, S. Y., L. Shah, Jolly, A.M. & Wylie, J. L.. (2007). Determinants of injection drug user (IDU) syringe sharing: the relationship between availability of syringes and risk network member characteristics in Winnipeg, Canada. *Addiction*, *102*(10), 1626-1635.
- Sherman, S. G. & Latkin, C. A. (2002a). Drug users' involvement in the drug economy: implications for harm reduction and HIV prevention programs. *Journal of Urban Health-Bulletin of the New York Academy of Medicine*, *79*, 266-277.
- Sherman, S. G. & Latkin, C. A. (2002b). Drug users' involvement in the drug economy: implications for harm reduction and HIV prevention programs. *Journal of Urban Health-Bulletin of the New York Academy of Medicine*, *79*, 266-277.
- Shrout, P. E. & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychological Methods*, *7*, 422-445.
- Stephenson, K. & Zelen, M. (1989). Rethinking centrality: methods and examples. *Social Networks*, *11*, 1-37.
- Suh, T., Mandell, W., Latkin, C., & Kim, J. (1997). Social network characteristics and injecting HIV-risk behaviors among street injection drug users. *Drug and Alcohol Dependence*, *47*, 137-143.
- Tobin, K. E., Hua, W., Costenbader, E. C., & Latkin, C. A. (2007). The association between change in social network characteristics and non-fatal overdose: results from the SHIELD study in Baltimore, MD, USA. *Drug and Alcohol Dependence*, *87*, 63-68.
- Tobin, K. E., M. Davey-Rothwell, & Latkin, C. (2010). Social-level correlates of shooting gallery attendance: a focus on networks and norms. *AIDS and Behavior*, *14*(5), 1142-1148.
- Unger, J. B., Kipke, M. D., De Rosa, C. J., Hyde, J., Ritt-Olson, A., & Montgomery, S. (2006). Needle-sharing among young IV drug users and their social network members: The influence of the injection partner's characteristics on HIV risk behavior. *Addictive Behaviors*, *31*(9), 1607-1618.
- Wallace, R. (1990). Urban desertification, public health and public order: 'planned shrinkage', violent death, substance abuse and AIDS in the Bronx. *Social Science and Medicine*, *31*, 801-813.
- Wasserman, S. & Faust, K. (1994). *Social Network Analysis Methods and Applications*. Cambridge: Cambridge University Press.
- Werb, D., Kerr, T., Li, K., Montaner, J., & Wood, E. (2008). Risks surrounding drug trade involvement among street-involved youth. *American Journal of Drug and Alcohol Abuse*, *34*, 810-820.
- Worthington, R. & Whittaker, T. (2006). Scale development research: a content analysis and recommendations for best practices. *Journal of Counseling Psychology*, *34*, 806-838.

# BRIEF REPORTS

## Social Conformity in Networks

---

**Ian McCulloh**

*Curtin University Business School  
School of Information Systems  
Perth, Western Australia*

### **Abstract**

Centrality in a social network is found to have a significant effect on Asch-type conformity. Friendship affinity and respect social network data was collected on two different groups of actors. The effects of Asch-type conformity were empirically tested on central actors and peripheral actors in each group using a culturally appropriate version of Asch's test. Findings show that central actors are less willing to conform and peripheral actors are more willing to conform than expected in Asch-type social conformity experiments.

### **Authors**

*Ian McCulloh, Ph.D.* is a visiting research fellow at the Centre for Organisational Analysis in the School of Information Systems at the Curtin University Business School in Perth, Western Australia.

### **Notes**

Thanks to Dominick Lombardi for his assistance in designing, executing and recording data for the empirical experiments.

*Please email all correspondence to Ian McCulloh at [cusum6@gmail.com](mailto:cusum6@gmail.com).*

## 1. Introduction

This article reports the empirical results of a series of experiments focused on understanding social network impacts on social conformity. Social conformity is the process of changing attitudes, behaviors, and beliefs to match group norms. Social networks are mathematical models of group relationships. It is hypothesized that individuals' position in the social network of a group may impact their pressure to conform to group norms and beliefs.

There is extensive literature investigating social conformity. The earliest studies of social conformity focused on ambiguous stimulus (Sherif, 1936). Solomon Asch (1952, 1955, 1956) conducted several experiments involving unambiguous stimuli. His experiments required subjects to report their observations of line length in a group where the other individuals would unanimously report wrong answers to selected questions. He found that 36.8 percent of subjects would deny their own observation and conform to the group answers. There is extensive literature reporting on the causes of conformity. (Abrams et al., 1990; Burnkrant & Cousineau, 1975; Cialdini, 2003; Cooper, 1979; Eagly, 1978; Eagly & Carli, 1981; Gerard, 1953; Linde & Patterson, 1964; Milgram et al., 1969; Turner, 1991; Williams & Sogon, 1984). Conformity experiments have been replicated across cultures (Askevis-Leherpeux & Zaleska, 1975; Avramov-Kiwetz & Game, 1974; Amir, 1984; Chandra, 1973; Huang & Harris, 1973; McKissack, 1971; Meade & Barnard, 1973; Neto, 1995; Rodrigues, 1982; Sistrunk & Clement, 1970; Sistrunk, Clement, & Guenther, 1971; Timaeus, 1968; Whittaker & Meade, 1967).

While Asch type conformity experiments have been extensively replicated, no experiment has specifically investigated the affect of social network position on conformity. Perhaps social network theories and centrality can explain variance in conformity among culture, gender, and group size. The remainder of this paper presents an Asch type conformity experiment conducted on two military groups. Social network data was collected on both groups. Subjects of the experiments were selected based on their network centrality within their group. While the experiments were limited in scope, results and conclusions are presented that provide compelling insight into the importance of social network position on social conformity.

## 2. Procedure

The affect of social network position on Asch-type conformity was empirically tested on two social groups. The first social group consisted of 20 soldiers in a U.S. Army Military Police (MP) platoon. All soldiers were enlisted and had served between one and ten years in the U.S. Army. Due to various conditions, this population was restricted in an actor's ability to develop a social circle outside the group of actors in the platoon.

The group's leader had collected social network data, recording friendship affinity and respect to better understand informal power within the platoon. The leader collected the data by passing out a list of all members of the platoon with a box labeled "Friendship" and a box labeled "Respect" next to each name. The respondents were given the following instructions,

"Place a check in the box labeled 'Friendship' next to those people who you consider to be a friend outside of normal working hours," and "Place a check in the box labeled 'Respect' next to people who you think are squared away." The term "squared away" is a common term used throughout the U.S. Military to describe someone who is highly proficient at their job and displays outstanding military bearing. The social network data were collected approximately six weeks prior to the conformity experiment so that soldiers would not associate the social network survey with the experiment. Figures 1 and 2 show the social networks for friendship affinity and respect, respectively.



Figure 1. Friendship Affinity Social Network for MP Platoon.

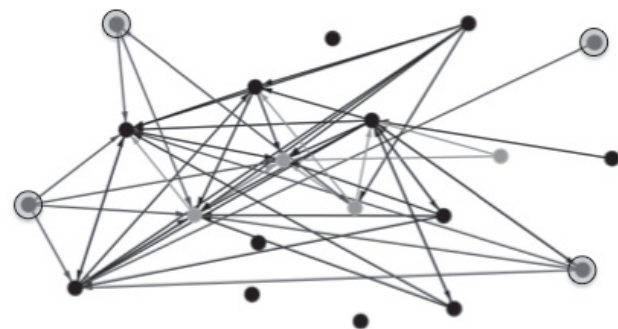


Figure 2. Respect "Squared Away" Network for MP Platoon.

An Asch-type conformity test was devised that would not appear abnormal to the respondents. Rather than asking respondents to report their observation of line-length, soldiers were asked questions commonly used for military promotion boards. In the U.S. Army, enlisted soldiers are promoted to the next rank if they have enough time in service and if they have earned a sufficient number of promotion points, which vary based on the soldier's career field (Headquarters, Department of the Army, 2010). The maximum number of points a soldier can earn are 800. Up to 150 points are awarded in a promotion board. A promotion board is convened by senior enlisted soldiers who will ask the soldier a variety of trivia questions that non-commissioned enlisted soldiers are expected to know. Questions may include knowledge of important military regulations, uniform appearance, general orders, first aid and many other topics. It is thus common for U.S. Army platoons to spend time practicing their knowledge in events that resemble trivia drills. Leaders may plan various games or other events to make

studying less monotonous or more effective. It is through this type of forum that the conformity test was delivered.

Forty questions were selected from a study guide for U.S. Army promotion boards (<http://www.armystudyguide.com>). Unlike most promotion boards, the investigators developed three multiple-choice answers for each question. The author's favorite question was, "When do you place a tourniquet on a neck wound?" Possible responses were "A) Never; B) Only if it is an artery or vein; C) If it is spurting blood". The answer is obviously A, since placing a tourniquet on a neck wound would kill the patient.

The subject of the experiment was selected based on their centrality in the social networks. The centrality values of betweenness, closeness, and in-degree centrality for both the friendship and respect networks were scaled between 0 and 1 and the six values were summed to create a composite aggregate centrality score. The four actors highest in aggregate centrality and the four actors lowest in aggregate centrality were selected to be respondents. Central actors are colored gray in Figures 1 and 2, while peripheral actors are colored redhighlighted with a shaded circle overlaying the node. There were two noncommissioned officers and two enlisted soldiers in each group, which allowed for a convenient control for military rank. The other actors were selected to be confederates of the experiment and provide incorrect responses to 30 out of 40 total questions. The seven respondents who were not participating in a current iteration of the conformity test were assigned other military details so they would have a legitimate reason to be away from the group conducting the conformity experiment. The respondent and confederates of the experiment sat in a conference room around a long table. The platoon leader provided the following instructions,

*"Today we will be preparing soldiers for the upcoming promotion board. In my psychology class, we have been studying memory and it is more difficult to learn answers to board questions when there are no clues or choices. I know that there are no multiple-choice answers in the board, but we are going to try preparing with multiple-choice questions to see if it helps soldiers learn answers better. This is how this will work: I am going to show you a question on a power point slide with three multiple-choice answers. I will read the questions and the answers out loud. You will then take turns answering the questions out loud. I will write down your answers. In order for me to test whether this approach works, it is important that you do not answer the question out of turn. You cannot make any comments about other soldiers' answers either. Are there any questions before we begin?"*

The platoon leader recorded the number of incorrect responses that the respondent gave in order to conform to the group. There were no questions where the respondent provided an incorrect response following correct responses from the confederates. In all cases where a respondent provided an answer that was different from the group, was for an incorrect response by the group. This may be due to the difficulty level of the ques-

tions asked. For two of the central actors, they were outspoken and questioned the confederates' intelligence based on their responses. The platoon leader had to keep reminding them to be quiet during the activity. The other two central actors looked visibly concerned by the incorrect responses to questions, but refrained from responding.

The experiment had several additional design features to ensure that it was conducted smoothly, especially since a single failure on a trial could potentially be communicated throughout the group and bias the entire experiment. The power point slides where soldiers were intended to provide an incorrect response had a slight variation in the logo that appeared in the upper left corner of the power point slide. This ensured that the confederates of the experiment knew when they were supposed to provide an incorrect response. The platoon leader conducted a rehearsal, going through all 40 questions with the confederates so that they understood how the experiment was to be conducted and gave them a chance to laugh at some of the more humorous incorrect responses prior to seeing it when the respondent did. The questions were randomized for each trial. The platoon leader was able to complete all trials on eight respondents within one day. None of the respondents had any contact with other respondents until after completion of the experiment.

The other group consisted of a platoon of 31 cadets in a military academy. The military academy provided the cadets an undergraduate college education and military training. Graduates earn a bachelor degree and a commission in the U.S. military. This platoon attended diverse classes and members pursued different academic majors throughout the day. They conducted military training on selected weekends and lived in the same dormitory. Promotion is based on the class year and to create leadership opportunities, therefore, rank is not permanent and promotion boards are not conducted. Social network data for friendship affinity and respect were collected on the cadet platoon approximately one week prior to the conformity experiment using the same protocol as for the MP platoon.

The cadet platoon leader administered the conformity test to the platoon using the same basic protocol and design as for the MP platoon. Unfortunately, cadets do not participate in promotion boards, nor does military trivia provide any benefit for advancement. Thus, it is highly uncommon for cadets to conduct an activity similar to the experiment. The instructions to respondents were altered slightly,

*"When you are commissioned your soldiers will have to prepare for promotion boards in order to advance. Good officers care about their soldiers. In my experimental psychology class we are studying memory. I am conducting a study to determine methods for increasing a soldier's ability to learn the military knowledge necessary for advancement. I am going to show you questions on power point slides with three multiple choice answers. I will read the questions out loud. You will then take turns answering the questions out loud. I will write down your answers. It is important that you do not answer questions out of turn to prevent bias in the experiment. It is also important that you do*

*not make any comments about other soldiers' answers. Are there any questions before we begin?"*

Figures 3 and 4 show the friendship and respect networks, respectively, for the cadet group. Gray colored nodes are central actors and highlighted-shaded nodes are the peripheral actors. Only five subjects were included in this series of experiments. The other intended subjects either became aware of the experiment or were unavailable at the last minute. In addition, four of the most central actors were unavailable to serve as subjects of the experiment.

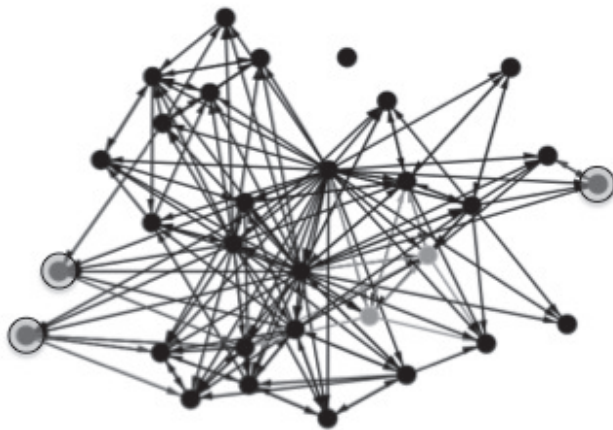


Figure 3. Friendship Affinity Social Network for Cadet Platoon.

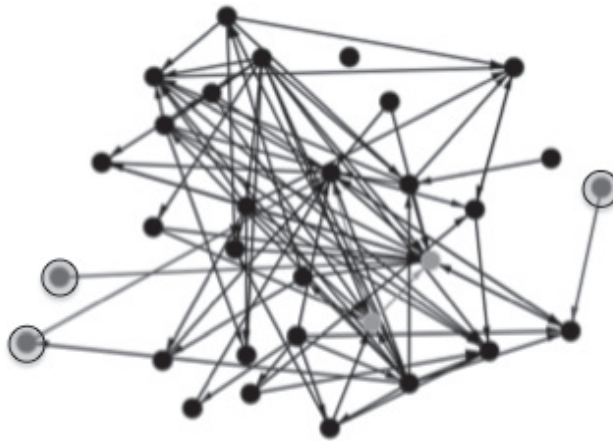


Figure 4. Respect "Squared Away" Network for Cadet Platoon.

Approval for these experiments was obtained by the appropriate U.S. Army Internal Review Board (IRB) for ethical treatment of subjects in a human experiment. Following the experiment, all subjects were debriefed on the experiment. It was important to review the questions where incorrect responses were given to make sure that soldiers were not misinformed regarding information that could affect their promotion and

possibly cause them to render improper first aid in an emergency. For the MP soldiers, their performance on the subsequent promotion board was monitored closely. Three subjects would not participate in a promotion board because they had either left the military, or successfully completed their last promotion board for advancement. Four subjects performed better on the following promotion board than they did on their previous promotion board at the earlier rank. One subject's performance on the promotion board was not observed.

None of the subjects or confederates were told that social network data was used in the experiment. They were not told that some actors were highly central to the group and others were peripheral. It was the opinion of the IRB that making respondents aware of their position in the platoon's social network may have adverse effects on their personal self esteem, especially if it had an effect on the subjects' responses.

Informal discussions with platoon members following the experiments revealed very different opinions on how successfully the experiments were performed. The MP platoon subjects did not suspect that they were participating in a conformity experiment. Two of the highly central subjects believed that something was unusual based on the responses of group members to certain questions, but they did not think that they were the subjects of an experiment. The cadet platoon, in contrast, knew that something was unusual almost immediately. They knew that their platoon leader was conducting the experiment primarily for a psychology class and that it had nothing to do with the military. Many suspected that it had something to do with the platoon's social network collected a week prior. Furthermore, all platoon members had taken an introductory course in psychology and had been exposed to Asch conformity in their respective course. Most of the participants were aware that the experiment was some version of Asch conformity.

Successful execution of this type of experiment depends upon subjects being unaware that they are in a conformity experiment. It is important to collect social network data at least a month prior to the conformity experiment to ensure that subjects and confederates do not link the two activities together. Furthermore, the questions must be embedded in a normal activity that actors would be expected to do. For these reasons, the MP platoon provided a better source of data for the experiment.

### 3. Results

The experiments conducted on the MP platoon were collected without incidence. The confederates of the experiment were very disciplined and professional and never let on that they were intentionally providing wrong answers. Results are reported in Table 1. The "N" in front of the subject identification code represents a noncommissioned officer, who would occupy a formal leadership position within the platoon. The "E" in front of the subject identification code represents an enlisted soldier, who would have no formal leadership responsibilities. The numbers of conforming responses to incorrect questions are reported in the second column. The remaining columns record the scaled network centrality values for the subjects.

There was a high correlation between network centrality and conformity in both groups. Table 2 displays the correlation between network centrality in the friendship affinity / respect networks and the number of questions where subjects conformed to the group to provide an incorrect response.

There is a higher correlation between betweenness centrality in the friendship network and conformity. This suggests that people who hold positions of informal power in the friendship network may feel free to speak their opinion without sanction from the social group. People who are peripheral to the group with betweenness centrality scores of 0, are more likely to conform to gain social acceptance or power within the group.

There is a higher correlation between conformity and centrality in the friendship network than centrality in the respect network. This may indicate that for these social groups, friendship is more important than respect within these social circles. It is also possible that performance in this particular task would not be perceived as a source of prestige and respect. For this reason, participants might be more affected by their friendship network than the respect network.

Conformity between the central and peripheral groups was also analyzed using a two-sample T-Test. For the MP platoon there was a statistically significant difference in conformity rate between the group of central actors and the group of peripheral actors ( $T = -3.23$ ,  $n = 8$ ,  $p = 0.0420$ ).

Results from the group of cadets were not as compelling. Informal interviews with subjects following experiments revealed that they all were aware that they were the subjects of a social psychology experiment. Additionally, most cadets, regardless of whether they were a subject or confederate of the experiment, were not familiar with correct responses to many of the questions. This may have increased conformity due to a lack of knowledge rather than social compliance. Finally, after five experiments the other intended subjects became aware of the study and the remaining planned experiments were unable to be completed. For the five subjects where successful data collection was completed, their conformity and scaled centrality scores are reported in Table 3.

There was a similar statistically significant difference in conformity for the cadet platoon ( $T = -3.54$ ,  $n = 5$ ,  $p = 0.0383$ ). Thus, even with the observed problems in data collection, results show that central actors were less likely to conform. There were a much higher number of conforming responses in the central group for the cadet platoon (18.5 conforming responses on average) compared to the central group for the MP platoon (2 conforming responses on average). There was no statistical difference between the number of conforming responses between the peripheral groups of the cadet and MP platoons, 24.6 and 22.5 respectively.

Table 4 reports the correlations between the number of conforming responses and the scaled network centrality measures for the cadet platoon. The correlations between network centrality and the number of conforming responses was much stronger in the cadet group than it was in the MP group. It is important to note, however, that in the cadet group, there is a high level of correlation among the subjects' centrality measures at 0.85, compared to 0.41 in the MP group. However, the fact that there remains high correlation between network position and

conformity reinforces the hypothesis that network centrality is an important variable in social conformity.

With the cadet group, in-degree centrality is more highly correlated with conformity than closeness or betweenness centrality. In the MP platoon the reverse is true. Furthermore, in the MP platoon, position in the friendship network appeared more significant than position in the respect network, whereas, in the cadet platoon, there was no significant difference between position in the friendship and respect networks. There are many possible reasons for this. There was a smaller sample size for the cadet group. The cadets had much more diverse social circles and greater opportunities for social connections. Finally, the cadets did not depend upon one another for their work performance. In a college setting, the performance of an individual does not directly affect the others in the group. If a cadet performs poorly in their academics, it does not affect the academics of his peers. Thus, a cadet's source of prestige in the respect network is essentially his popularity within the friendship network. For the MP platoon, however, job performance is largely separated from social performance and the friendship and respect networks have greater difference.

#### 4. Conclusion

This experiment shows empirical evidence of the impact of social network position on social conformity. Actors who are central to the group and have social acceptance are free to act in a manner inconsistent with the group and remain secure in their position. Peripheral actors who may be attempting to gain social acceptance have greater pressure to conform to the social group. Even in the cadet platoon, where subjects suspected they were in an experiment, there was a greater likelihood for peripheral actors to conform. This finding indicates that opinion leaders may have more freedom to deviate from group beliefs than they are constrained by their position in the organization.

There are several limitations to this study, however. None of the questions were necessarily related to cultural norms. Thus, a central actor's willingness to speak out against the group in this experiment may not remain consistent if he were violating a cultural norm. All of the questions were informational and did not have any ethical component. Therefore, this experiment does not address values or beliefs. It was not clear whether success in answering questions was even perceived as having value within the organizational culture of the group.

It is not clear how important the defined social group is to the subject's self-identity. For soldiers in the MP platoon, there are few enlisted soldiers on the installation that might provide opportunities to make friendships. Few of the soldiers have any extended family members in the area. This may create a greater need for social acceptance within the platoon. The cadets, in contrast, have relationships through their academic courses, sporting teams, extracurricular activities, and previous organizations. This provided the cadets much greater opportunity to find social relationships external to the platoon.

There is limited data. This experiment was conducted on two groups of military respondents. While there are significant findings of network position effect on conformity in both

**Table 1.** MP Group Conformity Responses and Network Centrality Measures.

Subject	Number of Conforming Responses	Between Centrality Friendship	Between Centrality Respect	Closeness Centrality Friendship	Closeness Centrality Respect	In-Degree Centrality Friendship	In-Degree Centrality Respect
N1	2	0.1213	0.0000	0.0884	0.0500	0.2632	0.5263
N2	0	0.1023	0.0139	0.0896	0.0625	0.2105	0.4737
E3	3	0.1754	0.0000	0.2021	0.0880	0.2105	0.0000
E4	3	0.1360	0.0288	0.2111	0.0969	0.0526	0.1053
E5	25	0.0000	0.0000	0.0500	0.0819	0.1579	0.0526
N6	10	0.0000	0.0019	0.0880	0.0950	0.0526	0.0526
N7	26	0.0000	0.0000	0.0848	0.0664	0.1053	0.0000
E8	30	0.0000	0.0000	0.0500	0.0657	0.0526	0.0000

**Table 2.** MP Group Correlations Between Network Centrality and Conformity.

	<i>Conformity</i>
Centrality Betweenness Friendship	-0.84
Centrality Closeness Friendship	-0.62
Centrality In Degree Respect	-0.61
Centrality In Degree Friendship	-0.48
Centrality Betweenness Respect	-0.48
Centrality Closeness Respect	-0.10

**Table 3.** Cadet Group Conformity Responses and Network Centrality Measures.

Subject	Number of Conforming Responses	Between Centrality Friendship	Between Centrality Respect	Closeness Centrality Friendship	Closeness Centrality Respect	In-Degree Centrality Friendship	In-Degree Centrality Respect
C1	17	0.0689	0.0039	0.3571	0.0333	0.2333	0.4000
C2	20	0.0151	0.0041	0.3614	0.0345	0.2000	0.2000
C3	23	0.0004	0.0000	0.2055	0.0344	0.1667	0.0333
C4	25	0.0142	0.0000	0.2055	0.0344	0.1667	0.0000
C5	26	0.0000	0.0000	0.2190	0.0356	0.1333	0.0000

**Table 4.** Cadet Group Correlations Between Network Centrality and Conformity.

	<i>Conformity</i>
Centrality Betweenness Friendship	-0.83
Centrality Closeness Friendship	-0.89
Centrality In Degree Respect	-0.97
Centrality In Degree Friendship	-0.97
Centrality Betweenness Respect	-0.90
Centrality Closeness Respect	-0.83

groups, more data is required for substantial findings. It is very difficult to obtain data of this nature, however. An investigator must obtain social network data on a group and then deliver a conformity test that is not obvious to the group members. This is very challenging as demonstrated with the suboptimal data collection for the cadet platoon. Fortunately, the protocol used in this experiment offers some important considerations for successful data collection as demonstrated with the MP platoon. This paper provides an important contribution in the design of network conformity experiments.

These experiments provide an important contribution in demonstrating a social network effect on conformity. The Asch conformity rate of 37% dropped to 5% when at least one other group member did not conform. While subjects reported good feelings toward the other non-conformist, they denied that person's impact on their own decision process. However, the significant difference in empirical findings suggests that other non-conformists play an important role in a person's decision to conform. This experiment provides structural context behind social conformity. Not only does network position offer an explanation for the impact of peers in social conformity, it provides a significant explanation of conformity in the first place. In the MP platoon, almost all of the central actors chose to answer according to their own views and did not conform. In contrast, almost all of the peripheral actors chose to conform. This trend continued in the cadet platoon, when the actors knew they were part of an experiment.

While this experiment was limited in size and scope, its potential findings are very important to understanding social conformity. With successful IRB approval, future experiments should repeat other Asch type experiments. What is the impact of another dissenting vote from a confederate of the experiment? Can culturally defined prestige variables be included in the experiment? What relationships matter most; friendship, advice, respect, or other relations? How many alternate social circles do actors have that can diversify their need for social acceptance within the group under study? All of these questions offer potentially better explanations for social conformity. This experiment demonstrates the importance of social network position in social conformity research. It is not the random dissenter or conformist that matters. Structure is a critical variable for conformity.

## References

- Abrams, D., Wetherell, M., Cochrane, S., Hogg, M. A., & Turner, J. C. (1990). Knowing what to think by knowing who you are: Self-categorisation and the nature of norm formation, conformity and group polarisation. *British Journal of Social Psychology, 29*, 97-119.
- Amir, T. (1984). The Asch conformity effect: A study in Kuwait. *Social Behavior and Personality, 12*, 187-190.
- The Army Study Guide. Retrieved September, 2008, from <http://www.armystudyguide.com>
- Asch, S. E. (1952). *Social Psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- Asch, S. E. (1955). Opinions and social pressure. *Scientific American, 193*, 33-35.
- Asch, S. E. (1956). Studies of independence and conformity. A minority of one against a unanimous majority. *Psychological Monographs, 70*(9, Whole No. 416).
- Askevis-Leherpeux, R., & Zaleska, M. (1975). Une approche experimentale de certains aspects du conformisme et de l'innovation dans les petits groupes [An experimental approach to certain aspects of conformity and innovation in small groups]. *Psychologie Francaise, 20*, 105-112.
- Avramov-Kiwetz, M., & Gaffie, B. (1974). Une approche experimentale de l'immunité à l'influence sociale dans les petits groupes [An experimental approach to resistance to social influence in small groups]. *Psychologie Francaise, 19*, 23-40.
- Burnkrant, R. E., & Cousineau, A. (1975). Informational and normative social influence in buyer behavior. *Journal of Consumer Research, 2*, 206-215.
- Chandra, S. (1973). The effects of group pressure in perception: A cross-cultural conformity study. *International Journal of Psychology, 8*, 37-39.
- Cialdini, R. B. (2003). Creating normative messages to protect the environment. *Current Directions in Psychological Science, 12*, 105-109.
- Cooper, H. M. (1979). Statistically combining independent studies: A meta-analysis of sex differences in conformity research. *Journal of Personality and Social Psychology, 37*, 131-146.
- Eagly, A. H. (1978). Sex differences in influenceability. *Psychological Bulletin, 85*, 86-116.
- Eagly, A. H., & Carli, L. L. (1981). Sex of researchers and sex-typed communications as determinants of sex differences in influenceability: A meta-analysis of social influence studies. *Psychological Bulletin, 90*, 1-20.
- Gerard, H. B. (1953). The effect of different dimensions of disagreement on the communication process in small groups. *Human Relations, 6*, 249-271.
- Headquarters, Department of the Army (2010) Army Regulation 600-8-19, Enlisted Promotions and Reductions. Washington, DC.: U.S. Army
- Huang, L. C., & Harris, M. B. (1973). Conformity in Chinese and Americans: A field experiment. *Journal of Cross-Cultural Psychology, 4*, 427-434.
- Linde, T. F., & Patterson, C. H. (1964). Influence of orthopedic disability on conformity behavior. *Journal of Abnormal and Social Psychology, 68*, 115-118.
- McKissack, I. J. (1971). Conformity in Ghana. *British Journal of Social and Clinical Psychology, 10*, 87.
- Meade, R. D., & Barnard, W. (1973). Conformity and anti conformity among Americans and Chinese. *Journal of Social Psychology, 89*, 15-25.
- Milgram, S; Bickman, L. & Berkowitz, L. (1969). Note on the drawing power of crowds of different size. *Journal of Personality and Social Psychology, 13*(2), 79-82.
- Neto, F. (1995). Conformity and independence revisited. *Social Behavior and Personality, 23*, 217-222.
- Rodrigues, A. (1982). Replication: A neglected type of research in social psychology. *Revista InterAmericana de Psicologia, 16*, 91-109.



- Sherif, M. (1936). *The Psychology of Social Norms*. New York, NY: Harper & Brothers.
- Sistrunk, F., & Clement, D. E. (1970). Cross-cultural comparisons of the conforming behavior of college students. *Journal of Social Psychology, 82*, 273-274.
- Sistrunk, F., Clement, D. E., & Guenther, Z. C. (1971). Developmental comparisons of conformity across two cultures. *Child Development, 42*, 1175-1185.
- Timaeus, E. (1968). Untersuchungen zum sogenannten konformen Verhalten [Studies on conformity behavior]. *Zeitschrift für Experimentelle und Angewandte Psychologie, 15*, 176-194.
- Turner, J. C. (1991). *Social Influence*. Milton Keynes, England: Open University Press.
- Whittaker, J. Q., & Meade, R. D. (1967). Social pressure in the modification and distortion of judgment: A cross-cultural study. *International Journal of Psychology, 2*, 109-113.
- Williams, T. P., & Sogon, S. (1984). Group composition and conforming behavior in Japanese students. *Japanese Psychological Research, 26*, 231-234.

# DEN

## Data Exchange Network

### The “AIRNET2000” Data and AIRNET Program

---

**Zachary Neal**  
*Michigan State University*  
*East Lansing, Michigan*

---

#### 1. Overview

The AIRNET2000 dataset was derived from the U.S. Bureau of Transportation Statistics’ Origin and Destination Survey, using the AIRNET program. It includes four longitudinal networks that describe passenger air traffic among 108 U.S. metropolitan areas annually, from 2000 – 2011 (12 waves). The route network describes all passenger movements between cities, while the origin-destination network describes passenger movements between an initial origin and final destination cities, omitting intermediate connections. The business and leisure networks are subsets of the origin-destination network, reflecting the origin-destination movements of passengers likely traveling for business or leisure, respectively. A distance matrix, recording cities’ great circle distances in kilometers, is also provided to facilitate spatial visualization and analysis.

The AIRNET2000 dataset represents a sample of the network data that can be generated using the AIRNET program, written for Stata. The program can be installed by typing ‘ssc install airnet’ in the Stata command line, and allows users to specify options to customize the resulting networks. Following installation, a detailed helpfile documenting use of the AIRNET program is available by typing ‘help airnet.’ Using the AIRNET program to produce network data provides access to:

- Wider longitudinal timeframes: From 1993 to present
- Narrower intervals: Quarterly, rather than annual
- Alternate nodes: US airports or US metropolitan areas
- Alternate formats: Edgelist or matrices

- Detailed decomposition of passengers’ routes by use of connections and layovers
- Alternate parameters for identifying likely business & leisure passengers

#### 2. Data Collection

The Origin and Destination Survey is conducted quarterly by the U.S. Bureau of Transportation Statistics, and provides details on a 10% random sample of all passenger air traffic in the United States. The raw data is provided in three nested data tables, each providing a different set of variables, and is available at: [www.transtats.bts.gov/Tables.asp?DB\\_ID=125](http://www.transtats.bts.gov/Tables.asp?DB_ID=125). The AIRNET2000 dataset was produced by executing the command “airnet, alpha(.05) minfare(20) maxfare(5000) metro(new) matrix distance” in Stata for each quarterly data release from 2000Q1 through 2011Q4, then aggregating the resulting networks annually. On a 2.5 GHz processor with 4 GB memory, this requires approximately 11 hours.

The basic networks reflect travel between all US airports, and are available in this form using the AIRNET program. The AIRNET program’s ‘metro’ option aggregates the 139 airports defined by the Federal Aviation Administration as “Primary Hubs” into US metropolitan areas. In the AIRNET2000 data, primary hub airports have been aggregated into the largest census-defined urban geographies – consolidated statistical areas (CSAs) and Metropolitan Statistical Areas (MSAs) using 2009 census definitions – to reflect that airports have large catchment areas.

### 3. Data Details

Response Rate	10% random sample, from 100% response
Non-Respondent Bias	n/a
Theoretical Grouping	Passenger air travel within the United States
Publications Using These Data	<ul style="list-style-type: none"> <li>• Neal Z. In press. Evolution of the business air travel network in the US from 1993 – 2011: A descriptive analysis using AIRNET. <i>Research in Transportation Business and Management</i>.</li> <li>• Neal Z. In press. AIRNET: a program for generating intercity networks. <i>Urban Studies</i>.</li> <li>• Neal, Z. 2013. “Types of Hub Cities and Their Effects on Urban Creative Economies.” In <i>Hub Cities and the Knowledge Economy: Seaports, Airports, Brainports</i>, Ashgate.</li> <li>• Neal, Z. 2013. <i>The Connected City: How Networks are Shaping the Modern Metropolis</i>. New York: Routledge.</li> <li>• Liu, X, Neal, Z, Derudder, B. 2012. City networks in the United States: A comparison of four models. <i>Environment and Planning A</i>, 44, 255 – 256.</li> <li>• Neal, Z. 2012. “Creative Employment and Jet Set Cities: Disentangling Causal Effects.” <i>Urban Studies</i> 49.</li> <li>• Neal, Z, Kassens, E. 2011. “The Business Passenger Niche: Comparing Legacy Carriers and Southwest During a National Recession.” <i>Journal of Air Traffic Management</i> 17: 231 – 232.</li> <li>• Neal, Z. 2011. “The Causal Relationship between Employment and Business Networks in US Cities.” <i>Journal of Urban Affairs</i> 33: 167 – 184.</li> <li>• Neal, Z. 2011. “From Central Places to Network Bases: A Transition in the US Urban Hierarchy, 1900 – 2000.” <i>City and Community</i> 10: 49 – 74.</li> <li>• Neal, Z. 2010. “Refining the Air Traffic Approach to City Networks.” <i>Urban Studies</i> 47: 2195 – 2215.</li> </ul>
Data Context	Derived from U.S. Bureau of Transportation Statistics’ Origin and Destination Survey
Respondents	Passenger air carriers operating in the US
Nodes	In AIRNET2000 Dataset: 108 US metropolitan areas Using AIRNET Program: US airports or US metropolitan areas
Edges	Airline passengers counts
Longitudinal	In AIRNET2000 Dataset: Yes, annual from 2000 – 2011 Using AIRNET Program: Yes, quarterly from 1993 – present
Temporality	n/a
Analytical or Pedagogical Utility	<p>Analytical –</p> <ul style="list-style-type: none"> <li>• Changes in cities’ accessibility &amp; air service adequacy</li> <li>• Route network: Diffusion of disease</li> <li>• Origin-Destination, Business, and Leisure Networks: Diffusion of information, wealth, etc.</li> </ul> <p>Pedagogical –</p> <ul style="list-style-type: none"> <li>• Demonstrating how valued graphs can be dichotomized</li> <li>• Demonstrating how directed graphs can be symmetrized</li> <li>• Illustrating spatial structure of graphs</li> <li>• Comparing flow networks carrying different resources</li> </ul>
Known Issues	None

#### 4. Data Files and Formats

The network data are provided as valued and directed adjacency matrices in four Excel workbooks, each containing 13 worksheets (tabs). The first 12 worksheets in each workbook contain a single year's network, while the final worksheet contains a distance matrix reflecting the great circle distance (in kilometers) between the metropolitan areas included in the network. The distance matrix can be used to obtain a geographically organized visualization of the networks and as a control variable in analyses.

**AIRNET2000R.xls** contains longitudinal route networks, annually from 2000 – 2011. Passenger movement in the route network is defined as a single take-off and landing. Each cell  $R_{ij}$  indicates the number of passengers who took off in city  $i$  and landed in city  $j$ .

Using the AIRNET program, and specifying the 'legtype' option, will decompose the route network passenger counts into four categories indicating the leg's position in the passenger's trip from initial origin to final destination: first, last, middle, or only. This decomposition is generally not useful for constructing separate networks, but is useful for measuring cities' roles as hubs (e.g. identifying the number of a city's total passengers that are connecting vs. terminal).

**AIRNET2000O.xls** contains origin-destination networks, annually from 2000 – 2011. Passenger movement in the route network is defined as from the initial origin city, to the final destination city, omitting any intermediate layovers or connections. Each cell  $O_{ij}$  indicates the number of passengers who started their trip in city  $i$  and ended it in city  $j$ .

**AIRNET2000B.xls** contains origin-destination networks for passengers likely traveling for business, annually from 2000 – 2011. Passengers likely traveling for business were identified using two criteria: (1) traveling alone and (2) paid a fare that was statistically significantly ( $\alpha = 0.05$ ) above the average fare for travel from the same origin and to the same destination in the same quarter. Each cell  $B_{ij}$  indicates the number of passengers meeting both criteria who started their trip in city  $i$  and ended it in city  $j$ .

**AIRNET2000L.xls** contains origin-destination networks for passengers likely traveling for leisure, annually from 2000 – 2011. Passengers likely traveling for leisure were identified using two criteria: (1) traveling with one or more companions and (2) paid a fare that was statistically significantly ( $\alpha = 0.05$ ) below the average fare from the same origin and to the same destination in the same quarter. Each cell  $L_{ij}$  indicates the number of passengers meeting both criteria who started their trip in city  $i$  and ended it in city  $j$ .

In the AIRNET2000 dataset business and leisure networks, statistical significance is assessed at  $\alpha = 0.05$ , and fares below \$20 (representing frequent flyer redemption fees) or above \$5000 (mostly representing private charter flights) were excluded from fare mean and standard deviation computations. Different values for these parameters can be specified using the AIRNET program's 'alpha,' 'minfare,' and 'maxfare' options.

All business and leisure networks generated by the AIRNET program, including those in the AIRNET2000 dataset, are restricted to passengers taking single-destination round-trips. More complex itineraries (e.g. a multi-city circuit) are relatively rare, and the complexity of their fare pricing makes the proxy identification of business and leisure passengers using these criteria impractical.

## Multi-Relational International Trade Networks, 1965-2000

---

**Matthew C. Mahutga**

*Department of Sociology  
University of California, Riverside*

---

### 1. Overview

Beginning with the seminal work of Snyder and Kick (1979), social network analysts have utilized trade relations to quantify the social structure of the world-economic system (e.g. Breiger, 1981; Clark, 2010; Clark & Beckfield, 2008; Kim & Shin, 2001; Mahutga, 2006; Mahutga & Smith, 2011; Nemeth & Smith, 1985; Smith & Nemeth, 1988; Smith & White, 1992). Some of this work utilizes data from the International Monetary Fund's Direction of Trade Statistics (IMF 2012). These data provide total trade for each country dyad, and cover a large number of dyads over a relatively long period of time (1950 to the present) (see Lloyd et al., 2009 for a review). One advantage to the Direction of Trade Statistics is that they are easy to work with because there are only  $N(N-1)$  data points in each year, and there is little year on year missing data because it is relatively easy for state agencies to accurately record total imports and exports with each of their partners. While there is much to learn from these data, total trade masks a significant amount of inter-industry variation in the structure of international trade (e.g. Hidalgo et al., 2007).

The other major data source is the United Nations Commodity Trade Database (UNCOMTRADE), which covers a large number of country dyads over a relatively long period of time (1962 to the present) (United Nations, 2012). A major advantage to the UNCOMTRADE data is that it disaggregates dyadic trade flows into industry and sub-industry categories, and thereby allows users to analyze inter-industry variation in the structure of international trade. These advantages present two unique challenges. First, while both sources are publically available, users face a more complex task to collect, clean and organize the UNCOMTRADE data insofar as the data points scale with both  $N$  and the number of relations (each year now has  $N(N-1)R$  data points, where  $R$  is the number of relations). And, the detailed classificatory schemes provided by the UN make it much more difficult for (especially poor) state agencies to comply with the reporting requirements consistently from year to year, resulting in a substantial amount of missing data on a year-on-year basis.

The data recorded here overcome some of the obstacles to employing UNCOMTRADE data because they record dyadic

trade among a constant set of 94 countries that together account for 96 to 99 % of world trade, cover multiple commodity relations and span a relatively long period of time. In particular, the trade matrices contain ordered dyadic trade flows reported in three time points (1965, 1980, and 2000). The 15 particular industries covered represent the 5 distinct commodity clusters identified by Smith and Nemeth (1985). The 45 matrices include a constant node-set of 94 countries in each year. Moreover, the data include roughly 33 percent more cases than the UNCOMTRADE database records for the three specific years, owing to a set of procedures that allowed me to infer missing trade between reporting and non-reporting countries and between non-reporting countries.

These data provide network analysts a rare opportunity to apply network methods to multi-relational international trade networks. These trade data are unique relative to other publicly available data insofar as they cover multiple commodity-trade relations, three time points spanning thirty five years, and include a large sample of countries representing all world-regions and levels of development. The data described in this article should facilitate the wider usage of multi-relational commodity-trade data because they require minimal processing prior to analysis, which has probably been the single largest obstacle to their usage thus far. In what follows I describe the industries covered, how the data were collected and reported, and discuss the procedures I followed to gather missing trade flows.

### 2. Data Collection

#### 2.1 Industries

The UNCOMTRADE data base has nine industrial classification systems for categorizing the types of goods traded between countries. The data described here are classified according to the Standard International Trade Classification (SITC) Rev. 1. While each of the nine classificatory schemes has its own advantages, the major advantage to SITC Rev. 1 is that it extends back to the first year that the UN began collecting data. Contrarily, the newer alternative schemes cover fewer years because commodities cannot be categorized "backward" in time once new schemes are developed.

The SITC Rev. 1 system classifies commodities under a multi-digit scheme that varies from total trade to hundreds of unique five-digit codes. Shorter digits imply a higher level of aggregation. For example, the one-digit code “7” is “Machinery and Transport equipment”, which subdivides into three unique two-digit codes: “71” is “Machinery, other than electric”, “72” is “Electrical machinery, apparatus and appliances”, and “73” is “Transport Equipment.” Each of these two-digit codes subdivides into greater specificity. For example, “72996” is “Electrical carbons.”

The data reported here were collected at the two-digit level, which capture recognizable industries. And, while the two-digit level certainly misses some of the finer disaggregation communicated by the longer codes, many countries do not report beyond the two-digit level for reasons related to administrative and resource burdens, or to preserve national secrets. Thus, the sum of smaller digit reports rarely equal the volume of trade reported at the two-digit level, but the sum of the two-digit level flows equals the one-digit level.

**Table 1.** UN Commodity Categories Classified in Relational Categories from Smith and Nemeth (1988).

---

<b>1) High Tech/Heavy Manufacturing</b>
58) Plastic Materials, Regenerated Cellulose and Artificial Resins
69) Manufactures of Metal
71) Machinery – nonelectrical
<b>2) Sophisticated Extractive</b>
25) Pulp and waste paper
34) Gas, natural and manufactured
64) Paper, paperboard, and manufactures thereof
<b>3) Simple Extractive</b>
04) Cereal and cereal preparations
22) Oil seeds, oil nuts and oil kernels
41) Animal oils and fats
<b>4) Low Wage/Light Manufactures</b>
83) Travel bags, handbags, and similar containers
84) Clothing
85) Footwear
<b>5) Animal Products and Byproducts</b>
01) Meat and meat preparations
02) Dairy products and bird’s eggs
29) Crude animal and vegetable materials

---

There are fifty five two-digit codes in the SITC Rev. 1 system (United Nations 1963). Table 1 reports the fifteen industries covered by the data described in this article. The industries in Table 1 were selected on the basis of Smith and Nemeth’s (1988) factor analysis, and were analyzed in part by Smith and White (1992), Mahutga (2006), and in full by Mahutga and Smith (2011). Smith and Nemeth’s factor analysis reduced the

55 two-digit matrices to five unique factors within which commodity matrices were highly correlated. Substantively, Smith and Nemeth’s factor analysis implied that the fifty five two-digit codes reduced to five broad categories, within which individual commodities were more or less interchangeable. A quick scan of the commodity clusters provides some intuition to their analysis. For example, the matrices for commodity codes 01 (“Meat and meat preparations”), 02 (“Dairy products and bird’s eggs”), and 29 (“Crude animal and vegetable materials”) were among a group of highly correlated trade matrices that clustered on a factor that Smith and Nemeth labeled “Animal Products and Byproducts”. Clearly, countries that for whatever reason—climate, geography, factor abundance, etc.—excel at the production and export (or conversely, do not excel and therefore import) of one type of animal product and byproduct, also excel at others.

## 2.2 Imports, Exports and Units of Measurement

In order to compile UNCOMTRADE data, the UN asks countries to report both their exports to and imports from each other country, which makes it possible to rely on either reported imports or reported exports to assemble a trade matrix. Exports and imports are very highly but imperfectly correlated. For example, the correlation of the vector of the US’s reported exports to its partners with the vector of the US’ reported imports from its partners will approach 1, but the value of the US’ reported import from Mexico on any given relation may not correspond exactly to the value of Mexico’s reported export to the US on the same relation. However, reported imports tend to be more accurate because of the care taken by state agencies to record imports precisely for the purpose of tariffs (Durand 1953). In general, I therefore rely on reported imports to assemble the trade matrices here. Thus, the vast majority of the cell entries in each  $N \times N$  commodity matrix represent country  $j$ ’s reported imports from country  $i$ , except as noted below. The dyadic trade flows in these matrices record the dollar amount of the given commodity group in thousands of current (i.e. not adjusted for inflation) US dollars.

## 2.3 Sample Selection and Missing Data

Dyadic trade flows on each of the fifteen commodity groups described above were collected for a constant panel of 94 countries in 1965, 1980 and 2000. The countries are reported in Table 2. However, only 63 of the 94 countries detailed in Table 2 reported trade (either imports or exports) in each of the three years. In order to increase the coverage above 63, I sampled as follows. I first included any country that reported in each year. I then included any country that reported trade flows in at least two of the three time periods, and used the following strategy to fill in missing data for each country that did not report in one of the years. I began by following StatCanada in utilizing “mirror flows” (i.e. reported exports to missing countries from non-missing countries), which left systematically missing data for the possible trade ties between non-reporting countries. In order to fill in the flows between countries that did not report in a given year, I used reported imports from a tem-

**Table 2.** Countries and Respective UN Codes.

UN Code	Country Name	UN Code	Country Name
012	Algeria	388	Jamaica
024	Angola	392	Japan
032	Argentina	400	Jordan
036	Australia	410	South Korea
040	Austria	414	Kuwait
048	Bahrain	434	Libya
052	Barbados	450	Madagascar
058	Belgium	454	Malawi
068	Bolivia	458	Malaysia
076	Brazil	466	Mali
096	Brunei Darussalam	470	Malta
120	Cameroon	480	Mauritius
124	Canada	484	Mexico
140	Central African Republic	504	Morocco
144	Sri Lanka	528	Netherlands
148	Chad	554	New Zealand
152	Chile	558	Nicaragua
156	China	562	Niger
170	Colombia	566	Nigeria
178	Congo	579	Norway
188	Costa Rica	586	Pakistan
196	Cyprus	590	Panama
200	Czechoslovakia	600	Paraguay
204	Benin	604	Peru
208	Denmark	608	Philippines
218	Ecuador	620	Portugal
222	El Salvador	634	Qatar
230	Ethiopia	682	Saudi Arabia
246	Finland	686	Senegal
251	France	702	Singapore
266	Gabon	724	Spain
270	Gambia	752	Sweden
276	Germany	757	Switzerland
288	Ghana	764	Thailand
300	Greece	768	Togo
320	Guatemala	780	Trinidad/Tobago
340	Honduras	788	Tunisia
344	Hong Kong	792	Turkey
348	Hungary	818	Egypt
352	Iceland	826	UK
356	India	841	USA
360	Indonesia	854	Burkina Faso
364	Iran	858	Uruguay
372	Ireland	862	Venezuela
376	Israel	882	Samoa
381	Italy	891	Yugoslavia
384	Côte d'Ivoire	894	Zambia

porally proximate year. The 31 countries for which I filled in missing data in this way are as follows:

- 1965: Algeria, Angola, Bahrain, Barbados, Czechoslovakia, China, Ethiopia, Gambia, Indonesia, Jamaica, Kuwait, Malawi, Mauritius, Poland, Qatar, Saudi Arabia, Trinidad/Tobago, Uruguay.
- 1980: Chad, Côte d'Ivoire, Iran, Nigeria, Romania, Zambia.
- 2000: Benin, Brunei, Cyprus, Central African Republic, Sri Lanka, Congo (Democratic Republic), Gabon.

Thus, 129,735 (or 33%) of the 393,390 dyads reported here

were obtained with the procedure for handling missing data outlined above. Finally, users will note that Table 2 lists both Czechoslovakia and Yugoslavia even though neither existed as independent states in 2000. The trade flows reported in 2000 for Czechoslovakia and Yugoslavia were obtained by aggregating the imports reported by the former Czechoslovakian and Yugoslavian republics.

In sum, the 94 countries appearing in this sample appear if either they reported imports in every year, or I could rely on a combination of “mirror flows” and temporally proximate flows between non-reporting countries for no more than one missing year. The full sample is representative of all world regions and accounts for between 96 and 99 percent of world trade, between 92 and 98 percent of world GDP, and roughly 80 percent of world population through time.

### 3. Data Files and Formats

The data appear in two formats—excel and UCINET. In each, the file names correspond to the year and commodity code of each relation. For example, y196501 is commodity code 01 (Meat and Meat Preparations) in the year 1965. The excel files do not contain labels, but the accompanying excel file titled “labels” lists both the UN country code and country name in the same order as the countries appear in the rows/columns of the data files. The UCINET files include the UN country codes on the rows and columns.

### 4. Data Details

**Table 3.** Data details.

Response Rate	N/A
Non-Respondent Bias	N/A
Theoretical Grounding	These data are relevant to questions about the organizational structure of manufacturing industries worldwide, as well as change in these organizational structures over time
Publications Using These Data	These data appear in part in Boyd et al. (2010); Mahutga (2006); (forthcoming); Smith and White (1992) and in full in Mahutga and Smith (2011)
Data Context	N/A
Respondents	N/A
Longitudinal	Yes, 15 relations in 1965, 1980 and 2000
Temporality	The valued dyads are measured in current US dollars
Analytical Utility	Any analytic context calling for comparisons of network structure across relations and over time
Known Issues	See description for procedures employed to handle missing data

## References

- Boyd, John P., Fitzgerald, William J., Mahutga, Matthew C., & Smith, David A. (2010). Computing Continuous Core/Periphery Structures for Social Relations Data with MINRES/SVD. *Social Networks*, 32, 125-137
- Breiger, Ronald (1981). Structures of Economic Interdependence Among Nations. In Peter M Blaue, & Robert K. Merton (Eds.) *Continuities in Structural Inquiry* (pp. 353-79). London, England: Sage.
- Clark, Rob (2010). World-System Mobility and Economic Growth, 1980 – 2000. *Social Forces*, 88, 1123-1152.
- Clark, Rob & Beckfield, Jason (2009). A New Trichotomous Measure of World-System Position Using the International Trade Network. *International Journal of Comparative Sociology*, 50, 5-38.
- Durand, Edward D. (1953). Country Classification. In Roy George, Douglas Allen, & J.E. Ely (Eds.) *International Trade Statistics* (pp. 117-29). Hoboken, NJ: John Wiley.
- Hidalgo, C.A, Klinger, B., Barabasi, A.L., & Hausmann, R. (2007, July). The Product Space Conditions the Development of Nations. *Science*, 317, 482-487.
- IMF (2012). *Direction of Trade Statistics*. Washington D.C: International Monetary Fund.
- Lloyd, Paulette, Mahutga, Matthew C., & de Leeuw, Jan (2009). Looking Back and Forging Ahead: Thirty Years of Social Network Research on the World-System. *Journal of World-System Research*, 15(1), 48-85.
- Mahutga, Matthew C. (2006). The Persistence of Structural Inequality? A Network Analysis of International Trade, 1965-2000. *Social Forces*, 84(4), 1863-89.
- \_\_\_\_\_. Forthcoming. “Global Models of Networked Organization, the Positional Power of Nations and Economic Development.” Review of International Political Economy.
- Mahutga, Matthew C., & Smith, David A. (2011). Globalization, The Structure of the World Economy and Economic Development. *Social Science Research*, 40(1), 257-72.
- Nemeth, Roger J. & Smith, David A. (1985). International Trade and World System Structure: A Multiple Network Analysis. *Review*, 8(4), 517-160.
- Smith, David A., & Nemeth, Roger J. (1988). Research Note: An Empirical Analysis of Commodity Exchange in the International Economy: 1965-1980. *International Studies Quarterly*, 32, 227-240.
- Smith, David A., & White, Douglas R. (1992). Structure and Dynamics of the Global Economy: Network Analysis of International Trade 1965-1980. *Social Forces*, 70(4), 857-893.
- Snyder, David, & Kick, Edward L. (1979). Structural Position in the World System and Economic Growth, 1955-1970: A Multiple-Network Analysis of Transnational Interactions. *American Journal of Sociology*, 84(5), 1096-1126.
- United Nations (2012). COMTRADE. New York, NY: United Nations
- \_\_\_\_\_. (1963). “Commodity Indexes for the Standard International Trade Classification, Revision 1.” Statistical Papers M:38(II). New York, NY: United Nations.









**Connections** publishes original empirical, theoretical, and methodological articles, as well as critical reviews dealing with applications of social network analysis. The research spans many disciplines and domains including Communication, Anthropology, Sociology, Psychology, Organizational Behavior, Knowledge Management, Marketing, Social Psychology, Political Science, Public Health, Policy, Medicine, Physics, Economics, Mathematics, and Computer Science. As the official journal for the International Network for Social Network Analysis, the emphasis of the publication is to reflect the ever-growing and continually expanding community of scholars using network analytic techniques. **Connections** also provides an outlet for sharing news about social network concepts and techniques and new tools for research.

