

**INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH AND POLICY  
COLUMBIA UNIVERSITY  
WORKING PAPERS**

**CHAINS OF AFFECTION:  
THE STRUCTURE OF ADOLESCENT ROMANTIC AND SEXUAL NETWORKS**

**Peter S. Bearman  
Institute for Social and Economic Research and Policy  
Columbia University**

**James Moody  
Department of Sociology  
Ohio State University**

**Katherine Stovel  
Department of Sociology  
University of Washington**

**May 2002**

**ISERP WORKING PAPER 02-04**

**ISERP**  
Institute for Social and Economic Research and Policy

---

Data for this paper are drawn from the National Longitudinal Study of Adolescent Health (Add Health), a program project designed by J. Richard Udry and Peter Bearman, and funded by a grant HD31921 from the National Institute of Child Health and Human Development to the Carolina Population Center, University of North Carolina at Chapel Hill, with cooperative funding participation by the following agencies: The National Cancer Institute; The National Institute of Alcohol Abuse and Alcoholism; the National Institute on Deafness and other Communication Disorders; the National Institute on Drug Abuse; the National Institute of General Medical Sciences; the National Institute of Mental health; the Office of AIDS Research, NIH; the Office of Director, NIH; The National Center for Health Statistics, Centers for Disease Control and Prevention, HHS; Office of Minority Health, Centers for Disease Control and prevention, HHS, Office of the Assistant Secretary for Planning and Evaluation, HHS; and the National Science Foundation. The authors thank Douglas White, Martina Morris, Mark Hancock, and J. Richard Udry for helpful comments on previous drafts of this paper.

Corresponding author. Peter S. Bearman. Institute for Social and Economic Research and Policy. 814 SIPA Building, Columbia University, New York NY 10027. Email: [psb17@columbia.edu](mailto:psb17@columbia.edu)

## **Abstract**

Recognizing the actual structure of sexual networks is critical for modeling the potential for disease transmission, if disease is spread via sexual contact. This article reports the structure of an adolescent romantic and sexual network in a population of over 800 adolescents residing in a mid-sized town in the mid-western United States. Precise images and measures of network characteristics are derived from reports of relationships that occurred over a period of eighteen months between 1993 and 1995. We compare the structural characteristics of the observed network to simulated networks governed by similar constraints on the distribution of ties, and show that the observed structures differ radically from the randomly generated networks. Specifically, we find that real sexual and romantic networks are characterized by much longer contact chains and far fewer cycles. We then identify a set of micro-mechanisms that generates networks with similar structural features to the observed network. Implications for disease transmission dynamics and social policy are explored.

**Keywords:** Sexual networks, disease diffusion, adolescents, networks, norms

## INTRODUCTION

The literature on AIDS and other sexually transmitted diseases (hereafter, STDs) is vast, with much attention paid to the biological development of disease, to treatment, to highly sophisticated modeling, to the impact of STDs on the medical infra-structure, and to the social psychological effects of diffusion and acquisition. What is largely overlooked is that STDs are of sociological interest because of the confluence of epidemiological factors and social structure that enable or constrain their rapid diffusion through some populations. This article reports the structure of an extensive adolescent romantic and sexual network and considers the implication of this network for the dynamics of disease diffusion. This enables us to see, for the first time, how sexual network structures -- which are the product of rules governing partner choice -- influence both the overall prevalence of disease and the determinants of individual risk.

As in politics, for sexually transmitted diseases, the past casts a long and burdensome shadow over the present. Only persons who carry infection are capable of infecting others, and so it follows that an individual's risk of acquiring a sexually transmitted disease is conditioned by the STD status of their sex partner. One step back, it follows that a partner's probability of infection is conditioned by their prior partners' STD status. Working a second step backwards, it follows that the same rules apply, as they do to a third and fourth step, and on and on, back into the murky, tangled, and largely invisible past of partners' past partners' past partners for however long the time-ordered chain of past fluid-exchange relationships may be. Limited reflection suggests that these distal partners of our current partner's prior partner(s) are likely unknown to us. But if the past is invisible, its shadow drapes the present and poses risk for the future. The past snakes forward through the tangled vines of often long-extinct sexual relationships. The structure of this viper's tangle is unknown, yet fundamental for assessing risk and for understanding the dynamics of STD diffusion writ large. In this article, we describe and account for the etiology of the structure of an adolescent sexual network.

Why should we care? Each year, in the United States alone, millions of individuals discover that they carry a STD. The two leading STDs, herpes and human papillary virus (HSV2 and HPV, respectively) are chronic and, although subject to limited palliative treatment, not curable. Adolescent STD acquisition rates continue to outpace those of all other groups, and AIDS is now a leading cause of death among persons 18-45 years of age.<sup>1</sup> The future does not look much brighter. The literature focuses on three main reasons for the gloomy outlook; taken together, these provide a partial account for the persistence of STDs among adolescents. First, one-half of all adolescents in the United States over 15 years old report having had intercourse (Resnick et al 1997), and a significant proportion of these adolescents are inconsistent in their use of condoms, therefore heightening risk of STD acquisition and transmission (Bearman and Brückner 1999). Furthermore, many adolescents who have not had intercourse are sexually active in a substantively meaningful (if technically ambiguous) way, and most do not use condoms during non-coital sex<sup>2</sup>. Second, the majority of adolescents with an STD have no idea that they are infected (Holmes et al 1999), consequently, they may fail to protect their partners even if they would prefer to do so, all things being equal. And third, relative to adults, adolescents tend to form romantic partnerships of short duration, on average only 15 months, but with a strong skew towards relationships of extremely short (less than 4 months) duration (Laumann, et al 1994). Most sex in adolescent relationships, if it takes place at all, occurs quickly, within the first two months (Bearman, Hillmann and Brückner 2001). This combination of short duration partnerships, inconsistent safe-sex practices, and incorrect assessment of STD status provides a partial account for the diffusion of STDs among the sexually active adolescent population. As fundamental is the role that sexual contact structures play in STD transmission dynamics.

---

<sup>1</sup> While evidencing significant local variability, the incidence rates for all of the major sexually transmitted diseases – chlamydia, herpes, gonorrhea, and HPV – within the adolescent population in the United States have increased dramatically in the past decade. Roughly 5% of all sexually active adolescents have acquired chlamydia or gonorrhea (Aral et al 1999). Among sexually active black adolescents, 25% are likely to be infected with herpes (CDC 2000), and probably 40-50% of all sexually active females have had a previous HPV infection, now known to account for most cases of adult cervical cancer (Holmes1999).

<sup>2</sup> Specifically, of adolescents who report that they are virgins (that is, have not had sexual intercourse) roughly one-third have had genital contact with a partner resulting in orgasm in the past year. Thus virginal status does not mean that adolescents are not engaging in behaviors that are free of risk for HIV or STD transmission, although the non-coital fluid-exchange behaviors they do engage in carry less risk for both partners than intercourse.

Given that the structure of sexual networks is critical for understanding STD diffusion dynamics, it is surprising that social scientists have almost no idea what they look like. The insight we have is limited to that provided by a handful of ego-centered network surveys (Morris and Kretzschmar 1995, 1997; Laumann et al 1994; Laumann and Youm 1998), and a scatter of snowball samples of populations at highest risk to HIV acquisition, specifically male homosexuals (Klovdahl 1985) and IV drug users (Rothenberg, Potterat, and Woodhouse 1996; Friedman et al 1997). While revealing some characteristics of the local fluid-exchange networks in which individuals are embedded, ego-centered and snowball methods provide an insufficient foundation for understanding the global sexual network structures that affect population-level disease risk.

Here, we describe partnership patterns and network structure for one population of adolescents residing in a mid-sized American town, thereby providing the first images and measurement of key structural characteristics of the extensive romantic and sexual networks through which sexually transmitted diseases may diffuse. Our approach involves a series of analyses. We first describe the models of sexual networks that are discussed in the existing literature on sexually transmitted diseases. We then report the structure of the complete network for which we have data. We consider both cross-sectional and temporal (viral and bacterial) views and discuss the extent to which the cross-sectional view underestimates the potential for disease diffusion. We then compare the observed network to randomly generated networks of similar size and degree. As it is possible that micro-level bias in partner selection (for example, a tendency for individuals with similar attributes or practices to form partnerships) could induce the network structure we observe, we consider the determinants of partnership choice and show that the observed structure is not a by-product of micro-level preferences for partnership attributes. We then consider how such a structure could emerge. We propose, and test, a simple behavioral rule that accounts for the structure we observe. Implications for public policy are considered in the conclusion.

To anticipate the main findings, we show that: i) none of the commonly accepted models of sexual networks correspond to the structure we observe; ii) partnership preference models induce incorrect structural representations; iii) a set of simple normative rules governing partnership selection

induces the structure we observe; and consequently, iv) current intervention efforts that focus on cores, rather than the general population, are poorly conceived.

#### MODELS OF DISEASE DIFFUSION

Modern models of infection disease diffusion focus on the basic reproductive rate,  $R_0$ , which is the number of new infections produced by an infected individual, over the duration of infectivity (Anderson and May, 1991). When  $R_0$  exceeds 1, a self-sustaining epidemic occurs, when  $R_0$  is below 1 the disease dies out. In most models of disease diffusion, the reproductive rate is a function of three parameters: the probability that an infected individual will infect a partner over the course of a partnership (B), the average duration of infectiousness (D), and the structure of contact across a population (C).

In this formulation, the critical sociological parameter is C, the structure of contacts. Many epidemiological models of disease diffusion assume extremely simple contact structures; the simplest models assume random mixing among all members of the population. Under random mixing, the number of new infections is easily calculated as the number of susceptibles times the number of infecteds times the proportion of contacts between susceptibles and infecteds that result in infection. The result of a random mixing model is the classical "S" shaped diffusion curve, where one observes a slow start, followed by exponential growth, and then a decline either from recovery or death (Sattenspiel 1990). One can think of random mixing as the statement: "people choose partners independent of their characteristics." For many diseases, random mixing captures the essential aspects of the diffusion process. The sneeze of a flu-ridden person on a transatlantic plane sends viral and bacterial material through the air, potentially infecting all of the passengers, though those sitting next to the sick person (likely, strangers, randomly assigned to their seat) are at greater risk. Although we may feel otherwise in our less gracious moments, we know that the airlines did not select us to sit next to a sneezer, and that he or she did not pick us (to sneeze on) because of our characteristics.

#### *Models of sexual contact networks*

For sexually transmitted diseases, however, random mixing is clearly a poor approximation of the underlying social reality. Here, partner-selection processes count. Though more complex, models that

explicitly consider bias in partner choice may more closely reflect the social and behavioral processes associated with the selection of sex partners. Models of preferred-mixing (Koopman et al 1989; Sattenspiel 1990; Jacquez, et al 1988; Hethcote 2000) assume disproportionately high levels of contact between individuals in the same group. For example, persons will prefer to have sex with those who are similar themselves on the basis of race, religiosity, sexual preferences, and so on.<sup>3</sup> “Leftover” contacts occur between people of different groups, proportional to the level of sexual activity of these groups. Depending on the specific mixing parameters employed, preferred mixing models predict wildly different levels and patterns of disease spread. How we conceptualize these parameters depends on what we think about the underlying network structures that they strive to summarize. Strangely, none of these models work with sexual network structure in a meaningful way, despite the fact that it is through sexual networks that diseases pass from one individual to another.<sup>4</sup>

However, systematic differences in the sexual network structures that govern patterns of direct and indirect contact can have striking implications for disease transmission (Morris 1997). Three images of sexual networks prevail in the literature. The first image is that of a *core*. The network image of a core of densely interconnected individuals passing infection to one another (and hence causing re-infection for treatable STDs) and then diffusing infection out to a less densely interconnected population, is shown in panel A of Figure 1. Here, circles represent individuals and lines represent fluid exchange relationships. In panel A, membership in the core is indicated by shading, with core members shaded black.

-----  
Figure 1 about here  
-----

Under the general diffusion model, infection is sustained in a core constituted by individuals with multiple partnerships whose partners also have multiple partnerships with others in the core. For many

---

<sup>3</sup> Preferred mixing models need not assume homogeneous mixing within groups. In many contexts, for example, observed partnerships evidence age-skewing, matching older males with younger females. The implications of skewed age matching for HIV diffusion are explored in Morris 1993.

<sup>4</sup> Preferred mixing models operate on persons classified by attribute rather than by structural position. The models we develop subsequently, identify position as the critical element, such that attributes of persons are substitutable across positions. But, in an abstract sense, they are also preferred mixing models.

STDs, such as gonorrhea, a core group of individuals with multiple partners (all tied together) have been shown to sustain reservoirs of infection (Hethcote and Yorke 1984; Hethcote and van den Driessche 2000; Hethcote and Van Ark 1987; Aral, et al 1999). These infection reservoirs push  $R_0$  in the core above 1, allowing infection to remain endemic.

Core structures are associated with particular, measurable network properties. In a core, it is likely that an individual's past partner is tied through a short chain of sexual contacts to his or her current or future partner. Thus if cores are observed in a population, cyclicity will be extremely high and the length of chains connecting two individuals in the population (geodesics) will be low.

While core-based models account for the diffusion of bacterial STDs, the imagery of the core fails to describe what is known about heterosexual transmission of HIV in third world countries. For example, when transmission occurs in networks associated with long-distance truck-drivers using prostitutes, many of the groups that constitute possible infection reservoirs (like prostitutes) are structurally disconnected and do not *directly* pass infection to one another. In these situations, infections are likely to spread through more complex switching structures (Garnett, et al 1996).

We represent such a switching structure, or *inverse core*, in panel B of Figure 1. In an inverse core, an infected population of persons pumps infection out to others but does not pass infection directly among themselves. For instance, prostitutes might be infected by previously infected johns and then pass infection on to other johns, though not to other prostitutes.

By labeling (coloring) individuals with respect to their role in the diffusion process (sex workers, sex customers) we gain insight into the social dynamics of partnerships. The key difference between a core and an inverse core stems from the social organization of sexual relations, since johns are more likely than other potential infecteds to spread infection to individuals not in the graph (specifically, their regular sex partners). Structurally, an inverse core is a bipartite graph. As in a core, an inverse core structure is associated with high cyclicity and low path distance between individuals, and since viruses are "less concerned" with the social details that occupy us, the two structures hold similar potential for disease diffusion.

A third model suggests that disease diffusion dynamics are best described by *bridging processes*, specifically bridging between largely disjoint populations of individuals (Aral 2000; Gorbach et al 2000; Morris et al 1996). Consider two populations of persons engaged in different behaviors (i.e., a high-risk and a low-risk population) linked by a few individuals whose behavior patterns cross the boundary between each world. A network of this type is shown in panel C of Figure 1. In fact, this graph expands on one of the bridging triads in the preceding figures, drawing attention to the fact that the focal triad is embedded in a macro-structure different than that implied by either the core or inverse core models.

All three graphs provide, in inchoate form, the foundation for preferred mixing models. Under the core and inverse core models, the underlying structure for mixing rests on in-group bias based on number of partners defined by risk status (or on some other attribute, like IV drug use, associated with risk status). In the bridging model, model parameters for cross-group contacts are estimated, and random mixing within groups is presumed. The value of these kinds of models is conditional on the underlying network structure that the parameterization of in- and out- group contact is intended to proxy. Obviously, if the structures one actually observes bear no relationship to these underlying models, we must radically reconsider the usefulness of the model. For our data, this is the situation.

Specifically, we observe a network structure that has the appearance of a *spanning tree*, a long chain of interconnections between individuals, stretching across a population like rural phone wires running from a long trunk line to each individual house (Hage and Harary 1996; 1983). In these networks, the global structure is defined by a graph with few cycles, low redundancy, and consequently very sparse overall density. The shortest distance between any two randomly selected individuals (geodesic) is significantly higher than that observed in either the core or inverse core structures.<sup>5</sup> The spanning tree structure is represented as panel D of Figure 1.

Random network processes do not produce spanning tree structures. Rather, they appear in contexts where formal or informal rules preclude the enactment of specific partnerships; consequently

they are most frequently observed in large and complex generalized exchange systems, for example the exchange of valuables in the Kula ring (Hage and Harary 1996; Schweizer and White 1998). In the language of kinship structures, spanning trees are the product of negative proscriptions: sets of rules about whom one *cannot* be in a relationship with.

As noted above, models of disease diffusion implicitly assume network structures that correspond to one the first three images in Figure 1, yet we have had absolutely no complete population data from which to conclude that any of these models are more appropriate. Fundamental at this point is to learn more about these networks and their structural parameters. The critical parameters are the distribution of network components (sets of interconnected persons), the presence or absence of cycles (closed loops that could generate a core if of sufficient size), mean geodesic length, density, and reach. In this article, we describe these characteristics of an observed romantic and sexual network in a population of adolescents. The network structure we find closely approximates a spanning-tree. Since such structures require rules restricting partnership acquisition processes, we focus on identifying a parsimonious rule that could produce the structure we observe. We simulate network processes under this rule and demonstrate that we can generate network graphs whose structural images closely correspond to what we observe empirically.

#### CONTEXT AND DATA

Data for this article are drawn from the wave 1 component of the National Longitudinal Study of Adolescent Health (hereafter, *Add Health*), a longitudinal nationally representative study of adolescents in grades 7-12. In 1994, in-school questionnaires were administered to ~90,000 students in 140 schools. Over 20,000 students were followed up, and interviewed in their homes in 1995. In two large (N=1000 and N=1800) and 12 small schools (N<250) *Add Health* attempted home interviews with all students enrolled in a high school or middle school. The two large schools were selected purposefully. The adolescent in-home interview was conducted using audio-CASI technology for all sensitive health status and health risk behavior questions. Adolescents listened to the questions through earphones and directly

---

<sup>5</sup> The relationship between a spanning tree and mean geodesics is not artifactual. A star-graph is technically a spanning tree, but

entered their responses into a computer, thereby eliminating interviewer or parental effects on their responses (Turner 1996). Adolescents were asked to identify their sexual and romantic partners by unique ID. Consequently, from the saturated field settings, we have almost complete sexual and romantic network data. These data provide the basis for this article<sup>6</sup>.

*Context: “Jefferson High”*

In this paper we report on data for one of the two large high schools in which we attempted in-home interviews with all students. This school, which we identify as “Jefferson High School”, is an almost all-white high school of roughly 1000 students located in a mid-sized mid-western town. Jefferson High is the only public high school in the town. The town, “Jefferson City” is over an hour away by car from the nearest large city. Jefferson City is surrounded by beautiful countryside, home to many agricultural enterprises. The town itself is working class, although there remain some vestiges of better times. At one period, the town served as a resort for city dwellers, drawing an annual influx of summer visitors. This is no longer the case, and many of the old resort properties show signs of decay. The community is densely settled. At the time of our fieldwork, students were reacting to the deaths of two girls killed in an automobile accident. Despite this, fieldwork proceeded exceptionally well. Interviewers wearing yellow *Add Health* buttons to identify themselves were frequently approached by adolescents who wondered when they would be asked to participate in the study<sup>7</sup>. Over the course of our interview period, over 80% of all students in the school completed in-home interviews. Over 90% of the students on the school roster participated in the in-school survey.

Table 1 reports on the characteristics of the 10-12th grade students in Jefferson High, providing a comparison of Jefferson High (column 1) with all high-schools in our sample (column 2), all disproportionately (>75%) white schools (column 3), high schools of comparable size (column 4), and

---

is associated with short geodesics.

<sup>6</sup> For design, see: Bearman, PS., Jones, J, and Udry, RJ. 1997. *The National Longitudinal Study of Adolescent Health: Research Design*. [www.cpc.und.edu/projects/addhealth/design.html](http://www.cpc.und.edu/projects/addhealth/design.html). and for details of the network design, see Bearman, PS, Moody J, Stovel K. *The Add Health Network Variable Codebook*. 1997.

<sup>7</sup> Adolescents were given \$20.00 in appreciation for completing the interview. Just before Mothers Day and the prom, many adolescents were eager to be interviewed.

finally, disproportionately white high schools of similar size (column 5). Additional information about these measures is provided in the Appendix.

-----  
Tables 1 about here  
-----

Standard statistical tests show that Jefferson High is similar to other schools in the United States across most of the comparison variables. There are some small differences between schools, though. Jefferson students earn lower grades, are suspended more, feel less attached to school, and come from poorer families than those at comparable schools. They are more likely than students in other high schools to have trouble paying attention, have lower self-esteem, pray more, have fewer expectations about college, and are more likely to have a permanent tattoo. Compared to other students in large disproportionately white schools, adolescents in Jefferson High are more likely to drink until they are drunk. In schools of comparable race and size, on average 30% of 10th-12th grade students smoke cigarettes regularly, whereas in Jefferson, 36% of all 10th to 12th graders smoke. Drug use is moderate, comparable to national norms. Somewhat more than half of all students report having had sex, a rate comparable to the national average, and only slightly higher than observed for schools similar with respect to race and size. Nevertheless, if Jefferson is not Middletown, it looks like an awful lot like it. The adolescents at Jefferson High are pretty normal. In describing the events of the past year, many students report that there is absolutely nothing to do in Jefferson. For fun, students like to drive to the outskirts of town and get drunk. Jefferson is a close-knit insular predominantly working-class community which offers few activities for its youth.

For our purposes, the relative isolation of the community is an important factor, significant for the patterns of romantic partnership and sexual partnership choices we observe. The context provides a conservative test for our model, for if redundant structures exist they are most likely to appear on island populations not permeated by the currents of larger more cosmopolitan settings.

### *Romantic partnerships and sexual partnerships*

During the in-home interview, adolescents were asked if they were in or had been involved in a 'special romantic relationship' at some point during the past 18 months. Adolescents in such relationships were asked to describe their three most recent relationships, including any current relationships, and to identify their relationship partners.<sup>8</sup> Adolescents were also asked to identify up to three individuals with whom they had a non-romantic sexual relationship in the past 18 months. A non-romantic sexual relationship was defined as a relationship involving sexual intercourse that the respondent did not identify as special and in which the partners did not kiss, hold hands, or say that they liked each other. A large number of sexual, non-romantic, relationships were reported. For the vast majority of reported partnerships, start and end dates for all romantic and non-romantic sexual partnerships were collected. Just under one-quarter of all Jefferson students had never been in a romantic or non-romantic sexual relationship.

After collecting detailed information about partnerships, respondents were asked if their partner attended their school (or the middle school that fed students into the high school). If their partners attended either school, respondents were asked to identify their partner by unique ID. By this process, from the 832 students who completed the in-home interview, we collected data on 535 partnerships. We use these nominations to generate images the structure of romantic and sexual networks. In-school nominations compose 51.2% of all romantic nominations made by Jefferson students, and 39.4% of all non-romantic sexual partnership nominations. They involve roughly 75% of all students who reported ever-having a romantic relationship. In comparison to our other large saturated field-setting, located in an ethnically heterogeneous metropolitan area, where only 11% of all partnership nominations are directed towards other students, Jefferson's students are romantically oriented towards the school community

---

<sup>8</sup> Adolescents who did not identify that they had a special relationship were asked if in any relationship over the past 18 months they had 'held hands, kissed, or told someone that they liked or loved them'. If an adolescent was in such a relationship, then they were asked to identify their partner and describe their relationship. Both self-identified and behavior-induced 'partnerships' could, but did not necessarily, involve sexual intercourse.

which provides the focal point (Feld 1981) for adolescent social and sexual relations. Given the relative isolation of the community, this orientation is expected.

In general, female adolescents tend to be involved with older male adolescents. We observe a similar pattern in Jefferson. Ninth grade girls tend to be in relationships with 9<sup>th</sup> and 10<sup>th</sup> grade boys, 10<sup>th</sup> grade girls with boys in the 10<sup>th</sup> and 11<sup>th</sup> grades, and so on. Among all partnerships involving Jefferson students, we observe a mean grade difference of .9, less than expected by chance distribution (1.23), but evidence of a female preference for older boys (or male preference for younger girls). Not all partners, sexual or romantic, are drawn from school. On average, out-of-school partners were 3.21 years older than the respondent at the start of their relationship, although we observe a pronounced skew in the age-difference distribution.<sup>9</sup> Many of these out-of-school partners attended Jefferson prior to our survey, and most live in the same neighborhood. Subsequently, we discuss the implications of these out of school partners for the sexual networks that we observe, but immediately below, we focus only on the in-school romantic and sexual network structure.

#### OBSERVED ROMANTIC AND SEXUAL NETWORK AT JEFFERSON HIGH

We first present the actual structure of the romantic and sexual network at Jefferson High. Figure 2 maps the complete set of network components among the 535 students involved in a romantic or sexual relationship with another student at Jefferson High.<sup>10</sup> Time is suppressed in this representation.

-----  
Figure 2 about here  
-----

A few simple components occur with some frequency in Jefferson High. For example, 126 students are involved in isolated dyadic relations, and therefore the simple dyadic structure occurs 63 times at Jefferson. Note, however, that far more than 126 students at Jefferson are only involved in one

---

<sup>9</sup> One out-of-school sexual partner of two girls was, for example, 39 years old. Aside from a few exceptions like this, most of the students involved in out of school relationships have slightly older partners than themselves.

<sup>10</sup> In Figure 2, and in all discussions presented here, all romantic and sexual relationship nominations linking students are included, whether or not the nomination from *i* to *j* was reciprocated with a nomination from *j* to *i*.

relationship. Many of the more complex components also include students with only one partner; however, the partners of these students have multiple partners.

Structures involving three students are also fairly prevalent at Jefferson. Triads composed of one male and two females occur twelve times, and triads composed of one female and two males occur nine times. All told, a total of 189 students at Jefferson (35% of the romantically active students) are embedded in sexual and romantic network components containing three or fewer students. There are very few components of intermediate size (4-15).

Most striking is the emergence of a large component involving 52% (283) of the romantically involved students at Jefferson. This large component involves the vast majority of individuals with multiple partners. It has numerous short branches (the two most distant individuals are 37 steps apart) and is characterized by the almost complete absence of cycles. In short, it is an almost perfect spanning tree.

The larger components, and certainly the largest component, are not likely to be visible to, or meaningful to, the students at Jefferson. They reflect relationships that may be long over, and they link individuals together in chains far too long to be the subject of even the most intense gossip and scrutiny. Nevertheless, these components are real. The largest component, especially, is a social fact -- an invisible yet consequential macro-structure that is the product of individual agency. These large components identify the boundaries of the worst case scenarios for potential disease diffusion. While one-third of all students are embedded in small disjoint components of size 2 or 3, more than 50% of the students at Jefferson are chained together through romantic and sexual relationships that may involve the exchange of fluids. Recall that there are numerous individuals at the end of the small branches that make up the large component who have only one partner. While these adolescents have only had one partner, their risk for a STD may be significantly greater than an individual with multiple partners who is embedded in a smaller, disjoint, component.

### *Temporal Unfolding*

Because time is suppressed in Figure 2, the indirect pathways connecting individuals through temporally ordered relationships cannot be seen, but they are consequential for disease transmission. If at

$T_1$ ,  $A$  and  $B$  are partners, and if at  $T_2$ ,  $B$  and  $C$  are partners, then a meaningful directed (from a viral or bacterial perspective) path between  $A$  and  $C$  exists. In contrast, the directed path between  $C$  and  $A$  is not meaningful. Figure 3 reports the meaningful indirect and direct ties for the major component. In contrast, consider Figure 4, which shows the fewest indirect relations possible on the same component. It is self-evident that both the cross-sectional view presented in Figure 2 and the minimum arrangement structure in Figure 4 radically understate the disease diffusion potential of the observed network. We return to this issue subsequently.

-----  
Figures 3 and 4 about here  
-----

#### GENERATING THE STRUCTURE: COMPARISON TO SIMULATED NETWORKS

Because no one has ever seen a complete structural mapping of a romantic/sexual network on one population, there is no existing baseline against which to evaluate what we observe. Obviously the observed network stands in clear contrast to the structures that implicitly rest behind most epidemiological modeling efforts, but this may simply be a product of chance distribution. To test this, we simulate 1000 networks of the same size and degree-- that is, we randomly simulate 1000 networks with the same number of nodes and the same overall distribution of number of partners--and consider where the network at Jefferson falls relative to the distribution of simulated networks.

-----  
Figure 5 about here  
-----

Figure 5 presents box-plots comparing the Jefferson High network to the simulated networks across the set of six network characteristics relevant for STD diffusion dynamics. The reported values for all network measures are standardized to have a mean of 0 and a standard deviation of 1. The crosshatch within each box plot reports the median value; the inter-quartile range is shaded. The value we observe for Jefferson High is indicated by a black circle. Across *all* of these basic measures of network structure, the sexual and romantic network in Jefferson is significantly different from the networks produced through random simulation.

The first dimension, *density at maximum reach*, assesses the extent to which the overall network is connected. Here, we measure the density of the network that arises when direct ties are inserted between all pairs of ever-reachable individuals.<sup>11</sup> Relative to the 1000 simulated networks, the Jefferson sexual network is extremely connected. This means that students at Jefferson are more likely to have partners at school who have other partners at school who have other partners at school. The alternative is many dyads, or other small groups of linked adolescents that are ultimately disjoint from the rest of the population, for example, a core disconnected from a set of smaller components. All things being equal, heightened connectivity is associated with more efficient disease spread. Our observed structure, therefore, poses greater risks to adolescents than those expected to arise simply from the chance distribution of partnerships.

Moving across Figure 5, we next consider *network centralization*, a measure of the inequality of the centrality of persons in the network. We calculate centralization using Bonicich's centrality algorithm with a negative beta value (Bonacich 1987). This parameterization gives extra weight to individuals who connect otherwise unconnected individuals and less weight to those whose bridges are redundant to other bridges in the network. Relative to the simulated samples, the Jefferson sexual network is highly centralized, suggesting that some actors may play a central role in linking disjoint clusters.

Continuing to move from left to right, the next box-plot reports the *mean geodesic length* of the Jefferson sexual network structure in comparison to the simulated networks. A geodesic is the shortest path between two connected persons in a network. Mean geodesic length is the mean length of the shortest path between every connected pair in the network. Large geodesics indicate a long chain-like structure, with few alternate paths directly connecting persons. In contrast, a network containing cycles, a core, many redundant ties, or star-like structures are likely to be associated with smaller mean geodesic lengths. Relative to the simulated samples, the sexual network at Jefferson High has very long mean

---

<sup>11</sup> Alternatively, one could represent this measure as the mean number reachable in a network. Likewise, from the same framework one can assess whether the maximum reach of the largest component, in our case involving roughly one-half of all students in the school is to be expected by chance. Relative to the simulated samples, the largest component is almost two standard deviations greater than expected by chance.

geodesics. This is the result of the extremely large component, and the overall absence of “short-cuts,” or redundant ties. With respect to STD diffusion, the absence of redundancy places pressure on the values of the B (probability of transmission given contact) and D (duration of infectiousness) parameters discussed previously. If B and D are low, spanning trees are inefficient structures for diffusion of STDs.

Overall connectivity in the observed network depends on long chains, and is consequently extremely fragile. It follows that as the mean geodesic length is large, that the *maximum geodesic length*, a measure that captures how difficult it is to reach the most distant pair of connected persons, will also be large. Relative to the simulated networks, the most distant pairs of connected individuals in Jefferson are quite distant from one another. They are likely to not know that they are involved in the same sexual web, which exists as a social fact beyond the reach of ordinary cognition.

In both the simulated and observed networks, every person in the romantic network can reach, by definition, at least one other person (their romantic relationship partner). In addition to direct relations, those persons not in disjoint dyads may also be able to “reach” others, through the relations of their direct partners. Extending this logic generates an individual-level measure of the number of “reachable” individuals in the network. We consider the *skew of the reach distribution* -- how unequally the number of reachable partners is distributed across the population. If most of the population were in isolated dyads, the distribution would show a strong positive skew, and less efficient pathways for disease diffusion. In contrast, a global network with a single large component would show a strong negative skew. This is the case in Jefferson. Skewed reach distributions are a trace of contact structures with heightened potential for disease spread.

Structurally critical to the Jefferson network is the pronounced absence of *cycles*. The absence of cycles guarantees that we are unable to observe a densely interconnected core functioning as a disease reservoir. In comparison to the simulation models, the observed sexual network structure is characterized by significantly fewer cycles than would be expected by chance. Consequently, models based on an assumption that the underlying network structures on a population at STD or HIV risk evidence a core or

inverse core pattern are not realistic. They are likely to yield under-estimation of the potential disease flow in contexts where treatment rates are low (or, alternatively, when asymptomatic cases are frequent).

### *Structural Fragility*

These simulations show that the observed sexual network in Jefferson differs markedly from those generated by chance. We observe a network structure dominated by an extraordinarily large component that connects more than half of all the students who are romantically and sexually active in the school. Such a component ties individuals together into long chains of potential infectivity. This purely negative picture with respect to disease diffusion is attenuated, however, by the absence of cycles in the large component. The absence of cycles eliminates redundant paths and creates structural fragility.

In structures like spanning trees, structural fragility exists because with the deletion of single edges (relationships), sub-graphs of the large component can become disconnected. Consider again the analogy to phone lines: a break in the major trunk line separates a single component into two disjoint components, and prevents calls from traveling from one component to the other. In such systems, protection against such failure is achieved by adding lines to build redundancy into the system. For sexual networks, redundant lines -- which create cycles -- provide the foundation for cores, the incubators of epidemics. Subtle changes in local network composition can thus have dynamic effects at the macro-level, potentially creating barriers to efficient disease diffusion by generating small disjoint components.

Technically, a component of a graph is a maximal connected sub-graph, that is, a sub-graph which cannot be made larger and still retain the property that there is a path between all pairs of nodes and that there is no path between a node in the component and a node not in the component. A bicomponent is a component that has the property that all nodes are connected by at least two different paths, and that the addition of a node requires that it is connected to two nodes in the sub-graph. Identifying cut-points that lie between two bicomponents allows for visual representation of how fragile the macro-structure is to disruption. Returning to the temporally ordered indirect ties of Figure 3, Figure 6 shows how the larger structure of indirect ties breaks into a set of smaller bicomponents when single

pathways between nodes are eliminated. While each of the smaller bicomponents is dense, in the absence of redundant ties the network is easily split into separate components, and hence, separate epidemics.

-----  
Figure 6 about here  
-----

#### MICRO-MECHANISMS IN PARTNER SELECTION

Because the spanning tree structure we observe is not likely to appear by chance, some set of dynamic processes governing partnership selection must account for it. In the next several sections, we focus on identifying these micro-mechanisms. It is obvious that when individuals choose partners they do not base their choice on its contribution to the global macro-structure. Put most starkly, adolescents don't account for their partner choice by saying, "by selecting this partner, I maximize the probability of inducing a spanning tree." First, they cannot see the global structure and second, they don't care about it. They may care, however, about the local structure in which the partnership is embedded, and about the attributes their potential partner has.

One possibility is that individual partnership selection, on the basis of preferences for specific characteristics (relative to their own characteristics), cumulates into a spanning tree. We first consider this idea. A second possibility is that there is a simple normative rule governing choice that, if followed by most individuals, necessarily induces a spanning tree. This is the solution we consider subsequently. We identify a rule that, if followed, would induce the precise macro-structure we observe. We provide reasons for thinking that this norm is followed, even if adolescents do not articulate it.

#### *Homophily in partnership experience*

Everyday experience, a cursory glance at personal advertisements in the classified section of any newspaper, a brief inquiry into the underlying logic of dating or matchmaking services, and a wide body of research, all suggest that individuals select partners on the basis of characteristics, and that persons tend to prefer partners who are similar to them. It is possible that an invisible hand governs the selection of partners, conditional on the underlying distribution of characteristics, such that a spanning tree structure emerges at the macro-level, even if this was not intended (or desired).

The simplest, and most straightforward, idea is that experienced partners prefer experienced partners, so that there is homophily on number of prior partners. This seems to be the case for adults; and it is this logic that gives rise to the idea of cores (Laumann et al 1994). Much of the difference between the observed characteristics of the Jefferson sexual and romantic network and the simulated networks appear to be the product of the number of isolated dyads we find in Jefferson. Consequently, it is possible that the single large component involving half of all students is a mathematical byproduct of homophily on one partnership characteristic -- the number of other partners an individual has had. If the majority of the individuals with only one partner are involved with individuals who also have only one partner, it follows that those with multiple partners are constrained to be involved with persons who have also had multiple partners. The catenation of these individuals should, all things being equal, generate large interconnected components similar to the single large component we observe.

To test this idea, we simulate 1000 networks with fixed size and degree distribution as before, remove 63 dyads (involving the 126 persons whose single partner has only a single partner), and prohibit the generation of isolated dyads. Our test focuses on the expected structural characteristics of global networks and their components when we remove all dyads. Adding this dimension has a stunning impact on the structure of our simulated networks. Specifically, we observe global network structures in which the mean size of the largest component is similar to the Jefferson network ( $n=283$ ). As suspected, homophily in partnership selection among less experienced partners (those with only a single romantic involvement) provides an efficient micro-mechanism for the generation of a large component.

To consider network features other than component size, we again compare the simulated networks with the Jefferson network across the set of network measures salient for disease diffusion previously discussed in Figure 5. These results are shown in Figure 7. Although we obtain a perfect match to the expected size of the largest component by controlling for the number of isolated dyads, the structural characteristics of the observed Jefferson network are still unusual relative to the simulated networks.

-----

Figure 7 about here

-----

As before, values are standardized, setting the mean at 0 with a standard deviation of 1. The crosshatch within each box plot reports the median value, and the inter-quartile range is shaded. Black circles indicate the value we observe for Jefferson. Across all six network measures, the structure of the Jefferson sexual network remains significantly different than expected, although less so than under the previous simulation, where we did not constrain isolated dyads. The slightly improved fit results from smaller variances in the simulated networks.

While the additional constraint produces components of the same size as we observe empirically, the internal structure of the large component in the simulated networks differs radically from what we observe in our data. In short, the simulated networks still appear more like cores than like spanning trees. The main differences between the simulated and real networks, in mean geodesic length, network centralization, reach, skew of maximum reach, are the product of the absence of cycles. Because we find many cycles of length 4 in the simulated networks, but few in Jefferson, we believe that there must be a normative prohibition against partnerships that involve the creation of short, cycles. We subsequently identify and model such a prohibition. First, however, we explore a set of competing hypotheses that propose that preferential partnership selection processes, beyond a simple homophily rule for inexperienced partners, induces the macro-structure we observe.

#### *Attribute-based selection preferences*

Preferential selection on partnership experience level provides an efficient foundation for generating large components in adolescent sexual networks, but fails to generate a spanning tree. It is possible, however, that selection on characteristics other than romantic experience induce the observed structure. While the number of attributes and behaviors that could provide a foundation for preferential partnership selection is infinite, we consider the level of homophily across a set of attributes and behaviors that might reasonably be expected to govern partnership formation among adolescents. Table 2 reports homophily on selected attributes within partnerships in Jefferson. Panel A considers ordinal variables, Panel B considers categorical variables. Columns 1 and 2 are presented as potential local

baselines: for the ordinal variables, we have calculated the mean difference between all potential cross-sex pairs in the school (column 1) and the mean difference between all potential cross-sex pairs involving romantically active students (column 2). Column 3 reports the mean difference between partners for all observed romantic pairs; column 4 reports the mean difference for current romantic partnerships only. For categorical variables we calculate the probability that members of a pair are in the same category. Using difference in means tests, we consider whether the extent of observed homophily differs from the column 2 baseline.

-----  
Table 2 about here  
-----

While we see some evidence of homophily in romantic partnerships, it is not overwhelming. Adolescents at Jefferson tend to select partners with similar socioeconomic status, grade point average, college plans, attachment to school, trouble in school, drinking behavior, IQ and grade, though their preferences on these dimensions are weak. Adolescents have a great deal of leeway in terms of selecting potential romantic partners.<sup>12</sup> With respect to categorical similarity (Panel B), partners tend to be similar in terms of religion, sexual experience and smoking. Evidently, students who smoke prefer other students who smoke. Alternatively, students who smoke induce smoking in their partners, perhaps because only smokers can tolerate kissing smokers<sup>13</sup>

*Building macro-structure from homophily on partner preferences: a p\* approach*

Some argue that it is possible to consider how the complex intercalation of partnership preferences with respect to attributes of individuals might cumulate into a global macro-structure using p\* modeling approaches (for details, see Anderson, Wasserman, and Crouch 1999; Wasserman and Pattison

---

<sup>12</sup> Current partnerships (not shown) are even less similar; in these relationships the only tendency is for students to be in the same grade.

<sup>13</sup> Obviously, some judgement is required here. Adolescents may select partners on the basis of unobserved characteristics (or unobserved to us) that vary across individuals in a completely unsystematic way. This is, in one sense, what the idea of romantic love suggests. Our strategy is to identify a set of characteristics that are observable, common, and have face validity as salient attributes. We consider homophily on these, and then simulate the global structure that would arise, should these elements provide the basis for choice.

1996). In contrast to our simulation-based evaluation, the  $p^*$  framework incorporates information about individuals and structural constraints simultaneously, and estimates the conditional likelihood of the observed network. However, meaningful interpretations of parameter estimates generated by this family of models have not been developed. Nevertheless, we follow standard practice and build models from a simple random graph to more complex structures. We include the four pair-wise attribute measures that provide the best model fits (grade in school, smoking status, popularity, and attractiveness) as well as structural change statistics associated with 2-paths, 2-stars, 3-stars, 3-paths and 4-cycles. To fit the model we construct a male-female dyad data set and estimate the probability of a tie conditional on both individual characteristics and the structural change statistics. The overall model fits are reported in panel A of Table 3; panel B reports the parameter estimates of the attribute mixing for model 12.

-----  
Table 3 about here  
-----

The strongest mixing effect can be seen with attractiveness and smoking. Similarly attractive students (or, similarly unattractive students) are disproportionately drawn into pairs; on the other hand, smokers and non-smokers don't pair up often. However, the structural coefficients are incoherent, which suggests that while individuals have preferences, almost any structure is available for their expression. The relative independence of preference to structure provides insight into why such preferences can exist in the first place. If preferences could find expression in only a single structure, most people would be without partners. The fluidity of structures relative to preferences allows all partnerships to be the product of choice, even if the choices and preferences that underlie them are, in a fundamental sense, completely arbitrary, or, alternatively, highly individuated.

#### UNCOVERING GOVERNING NORMS

The macro-level network structure we observe is neither the simple product of individual preferences for partners with particular attributes, nor produced by random assignment. We now seek to identify a norm capable of inducing this particular form. Earlier we showed that while spanning trees may be efficient for disease transmission, they are structurally fragile. Whereas our investigation of

structural fragility explored the consequences of removing relationships, we now consider the effects of *random re-wiring* of the network. That is, we purposefully introduce random "error" into the data structure. Since the error we introduce is random, it is insensitive to norms that may govern partner choice. Consider the possibility of an incest taboo that restricts available partners to those who are a culturally agreed upon distance from ego. By "re-wiring," some relations prohibited by this rule will be added to the network, and therefore some structural properties of the new graph may no longer match the original data structure. Structural features(s) of the re-wired graph that deviate from the original graph may help us identify behavioral rules that govern that specific parameter. Since we observe a spanning tree, re-wiring is likely to induce cycles, and hence cores. If this is the case, we know that an underlying rule prohibits the formation of cycles; and the problem is simply to identify it. We identify this norm, and show its effects in the following section.

To re-wire a graph, we begin with the empirically observed network, select 5% of the relationships at random, and randomly reassign them, conditional only on the observed degree distribution. In this way, we ensure that individuals with many partners continue to have many partners, and that individuals with few partners don't suddenly gain partners. Figure 8 reports the effect of re-wiring, based on generating 1000 re-wired graphs. As before, we present box-plots for six structural features of social networks -- each critical for the diffusion of STDs. The reported values for all network measures are standardized to have a mean of 0 and a standard deviation of 1. The crosshatch within each box plot reports the median value, the inter-quartile range is shaded. A black circle indicates the Jefferson value.

-----  
Figure 8 about here  
-----

Compared with our earlier simulations, the re-wired graphs are quite similar to the observed network. All of the network centralization measures are fit well, as are the reach measures. The difference between the observed number of components in Jefferson and those arising from the simulations is trivial. The only parameter that is fit poorly is the number of cycles. The simulated networks have more cycles

than observed in Jefferson. Re-wiring isolates a single structural feature we have to account for -- in this case, the absence of cycles. Thus the only puzzle is: why are they absent?

*The basis for a spanning tree structure: Unarticulated partnership prohibitions*

To explain why cycles are absent at Jefferson, we return to the theoretical relationship between spanning trees and negative proscriptions. What kinds of relationships are prohibited?. While normative prohibitions against partnerships between persons with particular attributes could exist, in populations of this size such rules are unlikely to provide enough constraint to generate a spanning tree. Therefore, the simple answer is that the prohibited relationships are those which induce cycles. In a predominantly heterosexual population, the *smallest cycle* has a length of 4. Consider four individuals, Bob, Carol, Ted, and Alice. Imagine that Bob and Carol were once partners, but that Carol left Bob for Ted. Further imagine that Ted and Alice were partners, but that Ted dumped Alice for Carol. Should Bob and Alice date? From Bob's perspective, Alice is his former partners' former partner, or the former "lover" of his former girlfriends' current lover. Alice looks at Bob with the same lens. Her former boyfriend is dating the girl who left Bob. These scenarios can be represented by a graph, as in Figure 9, where lines indicate a relationship between nodes (here, persons), yielding cycles of length 4. Using simulation, we can operationalize the normative rule that persons do not date the former (or current) partner of their former (or current) partner simply by prohibiting all cycles of length 4.

-----  
Figure 9 about here  
-----

Following the strategy developed earlier, we now simulate random networks with the following constraints: fixed degree distributions matching those observed at Jefferson, no isolated dyads, and a single parameter that prohibits cycles of length 4. The question is whether or not this model specification generates graphs with structural features similar to those observed in our data. As anticipated, these simulations generate components that, with respect to all of the structural parameters we consider, are statistically similar to the observed Jefferson sexual and romantic partnership network, as reported in

Figure 10. Moving from right to left across the graph, we note almost perfect correspondence between the observed Jefferson sexual network and the 1000 networks simulated under a 4-cycle prohibition.

-----  
Figure 10 about here  
-----

The correspondence is highlighted by visual inspection of some graphs of the largest components arising from the randomly generated networks induced under this model specification. Figure 11 (panels a-d) shows reports the structure of four large components, selected randomly. One immediately sees network structures that are strikingly similar to the structure we observe in Jefferson. Given fixed degree, the sufficient condition for generating a spanning tree is the prohibition against cycles of length four. Such a prohibition may operate in Jefferson.

-----  
Figure 11 about here  
-----

*Status dislocation and closeness*

Comparison of the internal structure of these random networks and the observed Jefferson network shows that they are essentially isomorphic. A single parameter – a prohibition against dating (from a female perspective) ones old boyfriends’ current girlfriends’ old boyfriend – accounts for the structure of the sexual network we observe. The prohibition works only on cycles of length four, and provides a parsimonious account for the etiology of a large-scale sexual network. Most adolescents would probably stare blankly at the researcher who asked (males): is there a prohibition in your school against being in a relationship with your old girlfriend’s current boyfriend’s old girlfriend? It is a mouthful, but it makes intuitive sense. From the perspective of either males or females, independent of the pattern of “rejection,” a relationship that completes a cycle of length four involves status loss, since the partnership is defined as a “seconds partnership.” In intact communities where observation of temporally proximate partnerships is possible, such as a close-knit high school, we should expect to see avoidance of relationships that induce cycles of length four. Such avoidance should not, however, extend to larger

cycles, since these involve relationships in the distant romantic past that cannot be systematically observed.

## DISCUSSION

Disease diffusion is widespread among adolescent populations. The standard models that epidemiologists have at their disposal to consider the dynamics of diffusion carry implicit ideas about the contact structure through which disease passes. These implicit ideas are associated with distinct structural features of sexual networks. The most critical feature is the idea of a core, which is associated with the presence of cycles in networks. In our data, we find that these key structural features are absent, and we propose a reason for their absence, namely a norm against short cycles. Using this norm, we generate networks that are structurally isomorphic to the one we observe, suggesting that in adolescent society -- where partner choice is salient for local status -- it seems reasonable to assume that such a rule operates.

Not all of the sexual partnerships in Jefferson are directed towards other students. It is possible that the spanning tree structure we observe is a by-product of missing data on the prior partnerships of the out of school partners. We consider this unlikely for three reasons. First, most adolescents with out of school partners do not have in-school partnerships. Second, the majority of non-romantic sexual partnerships are privatized -- that is, disembedded from the peer network. And third, most out of school partnerships are temporally subsequent to in-school partnerships. Consequently, their impact on the structure of the observed network is necessarily modest. Nonetheless, the scope conditions for this paper are implied in the central finding and the mechanism proposed to account for it. Specifically, the mechanism presumes that actors can watch each other, that they are capable of recording immediately prior partnerships, and that they are susceptible to collective assessment of their personal choices. One can only fear losing status in the eyes of others if the others are watching and if one cares about their assessments.

These same dynamic processes may be absent for adults who are embedded in worlds larger and more disjoint than adolescents. Consequently, adults are capable of segregating audiences across settings in which they are embedded (work, leisure, play, school, etc) and are thus less subject to the scrutiny and

sanctioning of their peers than are adolescents. While face may be lost by inducing cycles of length four, it is also possible that for many adults, such cycles could be induced without anyone ever knowing. This is unlikely in a high school, where much social energy is devoted to understanding who is going out with whom. Consequently, one would not expect to observe perfect spanning trees among adult populations.

Spanning trees are highly efficient structures for diffusion when there is no risk of a structural break, since the absence of redundant lines maximizes the reach at lowest density. Electric provision systems would be set up as spanning trees, for example, if service providers did not worry about failing to deliver power to some customers. It follows that for diseases with long infectivity duration, transmission is most effective under a spanning tree, net of structural breaks. Under conditions where structural breaks are likely, the fragility of spanning trees becomes a structural liability, and widespread diffusion (of electricity, disease, rumor or fashion) becomes unlikely (Watts, 2002).

Against this background, we might ask what kinds of interventions work most effectively to *induce* structural breaks in the sexual networks of adolescents. Here the answer is exceedingly simple. Assume that some proportion of actors who are “reached” through an intervention decide to change their behavior. Under core and inverse core structures, it matters enormously *which* actors are reached. Under a spanning tree structure it does not matter at all which actors are reached, just that some are, since the structure is similarly sensitive to a break (the non-consummation of a relationship) at any site in the graph. Consequently, from a structural perspective, interventions targeted to high-risk individuals may be inefficient at reducing the overall level of disease in the population: more efficient is broadcast intervention to all actors. Ironically, early HIV prevention strategies that utilized broadcast diffusion techniques may have been more effective than recent interventions focused on isolating those seen as high-risk (though these interventions may reduce risk of disease acquisition at the individual level).

Epidemiologists, unable to directly observe or measure the structure of sexual networks, have tended to latch onto a single idea: specifically, the idea that the number of partners matters for STD diffusion dynamics. If, as their models assume, the real contact structures can be approximated as cores or inverse cores, in-degree distributions are meaningful. But this is not always the case. By itself, this points

to the salience of the key sociological parameters governing disease diffusion, why the rate of STD acquisition has been so intense among adolescents in the past decade, and why US social policy, focused on high-risk individuals, has failed to stem the flood of new infections.

## REFERENCES

- Anderson, C. J., S. Wasserman, and B. Crouch. 1999. "A P\* Primer: Logit Models for Social Networks." *Social Networks* 21:37-66.
- Anderson, R. M. and R. M. May. 1991. *Infectious Diseases of Humans*. Oxford University Press.
- Aral, SO. 2000. "Behavioral aspects of sexually transmitted diseases: core groups and bridge populations." *Sexually Transmitted Diseases*. 27:327-328.
- Aral SO, Hughes JP, Stoner B, Whittington W, Handsfield HH, Anderson RM, Holmes KK. 1999. Sexual Mixing Patterns in the Spread of Gonococcal and Chlamydial Infections. *American Journal of Public Health*. 89 (6): 825-833.
- Aral SO, Holmes KK, Padian NS, Cates W. 1996. Individual and Population Approaches to the Epidemiology and Prevention of Sexually Transmitted Diseases and Human Immunodeficiency Virus Infection - Overview. *Journal of Infectious Diseases*. 174: S127-S133 Suppl. 2.
- Bearman, P. *Aids and Sociology*. 1992. Primis. McGraw Hill. New York.
- Bearman, PS, Moody J, Stovel K. *The Add Health Network Variable Codebook*. 1997.
- Bearman PS, Jones J, Udry RJ. The National Longitudinal Study of Adolescent Health: Research Design. Carolina Population Center. University of North Carolina at Chapel Hill; 1997. Available at: <http://www.cpc.unc.edu/projects/addhealth/design.html>.
- Bearman PS and H. Brückner. 1999. *Peer Effects on Adolescent Girls' Sexual Debut and Pregnancy*. Washington D.C. National Campaign to Prevent Teen Pregnancy.
- Bearman, PS, H. Hillmann, and H. Brückner. 2001. PAIRS: Adolescent Relationship Dynamics. (Manuscript). Institute for Social and Economic Research and Policy, Columbia University.
- Bonacich, P. 1987. "Power and Centrality: A Family of Measures." *American Journal of Sociology* 92:1170-1182.
- Centers for Disease Control and Prevention. 2000. *Tracking the Hidden Epidemics: Trends in STDs in the United States*. Atlanta, Georgia.
- Feld, S. L. 1981. "The Focused Organization of Social Ties." *American Journal of Sociology* 86:1015-35.
- Friedman, S. R., A. Neaigus, B. Jose, R. Curtis, M. Goldstein, G. Ildefonso, R. B. Rothenberg, and D. C. Des Jarlais. 1997. "Sociometric Risk Networks and HIV Risk." *American Journal of Public Health* 87:1289-96.
- Garnett GP, Hughes JP, Anderson RM, Stoner BP, Aral SO, Whittington WL, Handsfield HH, Holmes KK. 1996. Sexual Mixing Patterns of Patients Attending Sexually Transmitted Disease Clinics. *Sexually Transmitted Diseases*. 23 (3): 248-257
- Gorbach PM, Sopheab H, Phalla T, Leng HB, Mills S, Bennett A, Holmes KK. 2000. "Sexual bridging by Cambodian men: potential importance for general population spread of STD and HIV epidemics." *Sexually Trasmitted Disease*. 27:320-6.

- Hage, P. and F. Harary. 1983. *Structural Models in Anthropology*. Cambridge: Cambridge University Press.
- Hage, P. and F. Harary. 1996. *Island Networks: Communication, Kinship, and Classification Structure in Oceania*. New York: Cambridge University Press.
- Hethcote, H. W. and J. W. Van Ark. 1987. "Epidemiological Models for Heterogeneous Populations: Proportionate Mixing, Parameter Estimation, and Immunization Programs." *Mathematical Biosciences* 84:85-118.
- Hethcote, H. W. and P. van den Driessche. 2000. "Two SIS Epidemiologic Models with Delays." *Journal of Mathematical Biology* 40:3-26.
- Hethcote, H. and J. A. Yorke. 1984. "Gonorrhea Modeling: A Comparison of Central Methods." *Math Biosci* 58:93-109.
- Hethcote, H. 2000. "The Mathematics of Infectious Diseases." *SIAM Review* 42:599-653.
- Holmes KK, Mardh PA, Sparling PF, Lemon SM, Stamm WE, Piot P, Wasserheit JN (Eds). 1999. *Sexually Transmitted Diseases*. 3rd ed. New York: McGraw Hill.
- Jacquez, J. A., C. P. Simon, J. Koopman, L. Sattenspiel, and T. Perry. 1988. "Modeling and Analyzing HIV Transmission: The Effect of Contact Patterns." *Mathematical Biosciences* 92:119-99.
- Klovdahl, A. S. 1985. "Social Networks and the Spread of Infectious Diseases: The AIDS Example." *Social Science Medicine* 21:1203-16.
- Koopman, J., C. P. Simon, J. A. Jacquez, and T. S. Park. 1989. "Selective Contact Within Structured Mixing With an Application to HIV Transmission Risk From Oral and Anal Sex." Pp. 316-49 in *Mathematical and Statistical Approaches to AIDS Epidemiology*, Editor C. Castillo-Chavez. Berlin: Springer-Verlag.
- Laumann, E. O., J. H. Gagnon, R. T. Michael, and S. Michaels. 1994. *The Social Organization of Sexuality: Sexual Practices the United States*. Chicago: University of Chicago Press.
- Laumann, E. O. and Y. Youm. 1999. "Racial/Ethnic Differences in the Prevalence of Sexually Transmitted Diseases in the United States: A Network Explanation." *Sexually Transmitted Diseases* 26:250-261.
- Morris, M. 1993. "Epidemiology and Social Networks: Modeling Structured Diffusion." *Sociological Methods and Research* 22:99-126.
- Morris, M. 1997. "Sexual Networks and HIV." *AIDS 97: Year in Review* 11(Suppl A):S209-S216.
- Morris, M. and M. Kretzschmar. 1995. "Concurrent Partnerships and Transmission Dynamics in Networks." *Social Networks* 17(Special Issue. Social Networks and Infectious Disease: HIV/AIDS):299-318.
- Morris, M. and M. Kretzschmar. 1997. "Concurrent Partnerships and the Spread of HIV." *AIDS* 11:641-48.

Morris, M., C. Podhisita, M. Wawer, and M. Handcock. 1996. "Bridge populations in the spread of HIV/AIDS in Thailand." *AIDS* 10:1265-71.

Resnick, M. D., P. Bearman, R. W. Blum, K. E. Bauman, K. M. Harris, Jo. Jones, J. Tabor, T. Beuhring, R. Sieving, M. Shew, M. Ireland, L. Bearinger, and J. R. Udry. 1997. "Protecting Adolescents From Harm: Findings From the National Longitudinal Study on Adolescent Health." *Journal of the American Medical Association* 9.10:832-43.

Rothenberg, R. B., J. J. Potterat, and D. E. Woodhouse. 1996. "Personal Risk Taking and the Spread of Disease: Beyond Core Groups." *Journal of Infectious Diseases* 174(Suppl. 2):S144-49.

Sattenspiel, L. Modeling the Spread of Infectious Disease in Human Populations. 1990. *Yearbook of Physical Anthropology*. 32:245-276.

Schweizer, T. and D. R. White. 1998. *Kinship, Networks and Exchange*. Cambridge: Cambridge University Press.

Wasserman, S. and P. Pattison. 1996. "Logit Models and Logistic Regressions for Social Networks: I. An Introduction to Markov Graphs and P\*." *Psychometrika* 61:401-25.

Watts, D. *Six Degrees: The Science of a Connected Age*. 2003. New York, WW. Norton. (Forthcoming).

Table 1. School Level Comparisons: Jefferson High and Other High Schools in the *Add Health* Sample

Variable	Jefferson High	All High Schools	Mainly White HS	HS 600-1000 Students	White HS, 600-1000 Students
Family SES	5.59	5.80 (1.17)	5.94 (.91)	5.73 (1.19)	5.80 (.82)
Percent In Poverty	.13	.16 (.14)	.09 (.09)	.16 (.14)	.12 (.12)
Log(Family Income)	3.61	3.55 (.42)	3.64 (.29)	3.54 (.47)	3.56 (.33)
GPA	2.48	2.83 (.26)	2.89 (.21)	2.77 (.21)	2.77 (.18)
Expect College Graduation	3.77	4.09 (.48)	4.16 (.36)	4.08 (.30)	4.03 (.24)
School Attachment	3.41	3.78 (.19)	3.73 (.25)	3.64 (.17)	3.58 (.13)
Trouble in School	1.41	1.23 (.13)	1.27 (.09)	1.26 (.11)	1.27 (.10)
Drunk	1.23	.88 (.33)	.94 (.32)	.92 (.31)	1.00 (.22)
Delinquency	.29	.26 (.07)	.25 (.06)	.27 (.05)	.25 (.05)
Hours Watch TV	10.29	13.73 (3.52)	11.78 (2.67)	13.71 (4.24)	11.90 (2.20)
Religiosity (Praying)	2.71	2.10 (.42)	2.22 (.48)	2.20 (.38)	2.35 (.35)
In-Degree	5.74	4.52 (.68)	4.91 (.79)	4.76 (.89)	5.04 (.57)
Self Esteem	2.89	3.03 (.09)	3.04 (.08)	3.02 (.09)	3.01 (.09)
Sexually Active	.59	.62 (.39)	.52 (.17)	.55 (.17)	.53 (.16)
Autonomy	6.02	5.50 (.43)	5.59 (.38)	5.59 (.36)	5.67 (.30)
Expect to get AIDS	1.52	1.52 (.14)	1.53 (.11)	1.57 (.09)	1.58 (.08)
Marry by 25	3.06	3.18 (.25)	3.31 (.20)	3.13 (.17)	3.19 (.12)
Attractiveness	3.48	3.57 (.21)	3.59 (.17)	3.57 (.21)	3.57 (.20)
AH_PVT	105.32	100.91 (8.83)	104.92 (2.92)	101.84 (6.47)	104.56 (3.04)
Two Biological Parents	.46	.53 (.13)	.59 (.10)	.54 (.14)	.56 (.09)
Smokes Regularly	.36	.25 (.11)	.30 (.09)	.27 (.08)	.30 (.06)
School Suspension	.40	.27 (.14)	.23 (.12)	.30 (.14)	.27 (.11)
Tattoo	.096	.06 (.04)	.05 (.03)	.06 (.03)	.06 (.03)
Number:	1	76	30	24	11

Mean, Standard Deviation in Parentheses (10<sup>th</sup>, 11<sup>th</sup>, and 12<sup>th</sup> grade students only)

Table 2: Homophily between partners

Panel A. Partnership Comparisons: Pair-Wise Difference in Ordinal Attributes  
 Cell Values are mean absolute difference between members of a pair

	Mean Difference: Entire School	Mean Difference: Romantically Active Students	Observed Distance: All Romantic Pairs
SES	2.54	2.52	2.23**
Grade	1.233	1.213	.839**
Grade Point Av.	.887	.866	.754**
Expect College Grad.	1.41	1.30	1.09**
School Attachment	.996	.977	.83**
Trouble In School	.789	.782	.746
Gets Drunk	1.48	1.48	1.25**
Delinquency	.391	.387	.372
Hours Watch TV (ln)	.96	.96	.98
Religiosity (praying)	1.51	1.53	1.51
In-degree	4.35	4.636	4.87
Self Esteem	.659	.673	.648
Autonomy	1.377	1.343	1.37
Expect to get AIDS	.697	.693	.691
Expect Married by 25	1.21	1.20	1.17
Attractiveness	.981	.968	.908
AH_PVT	13.14	12.466	10.55**

\* =  $p < .05$ , \*\* =  $p < .01$  ( $H_0$ : Observed Pair = Romantically Active Mean)

Panel B. Partnership Comparisons: Pair-wise similarity in Categorical Attributes  
 Cell values are probability that pair members are in the same category

	Cross-Sex Random Mixing: Entire School	Cross Sex Random Mixing: Romantically Active Students	Observed In- Group Association Rate: All Romantic Pairs
Religion	.329	.346	.38*
Sexually Active	.500	.499	.63**
Smoking	.557	.554	.63**
School Suspension	.544	.512	.582

\* =  $p < .05$ , \*\* =  $p < .01$  (Test is Chi Square for pair 1 by pair 2 classification table)

Table 3. p\* Models of Observed Romantic and Sexual Network

Panel A: Model Fits

Model	Model Description	# of parameters	Correlation between observed & predicted	-2 log likelihood score
1	Simple random graph model	1	NA	5765.21
2	fixed row column effects	446	.056	5533.38
3	2 + 2-stars	3	.232	5374.91
4	3 + 3-stars	5	.409	4617.08
5	4 + 3-path	6	.410	4617.03
6	5 + isolated dyads	7	.525	4511.97
7	1, limited to all non-missing data	1	NA	4275.61
8	7 + 2-stars (= 3)	3	.224	4012.52
9	8 + 3-starts (= 4)	5	.401	3487.85
10	9 + 3-paths (= 5)	6	.401	3487.53
11	10 + isolated dyads (=7)	7	.482	3443.27
12	11 + attributes	11	.500	3390.79

Panel B: p\* parameter estimates for model of preferential partnership choice, model 12

Variable	Estimate
Density	1.020 <sup>***</sup>
2-stars (M→F)	-2.477 <sup>***</sup>
2-stars (F→M)	-3.864 <sup>***</sup>
3-stars (M→F)	0.910 <sup>***</sup>
3-stars (F→M)	1.679 <sup>***</sup>
3-Paths	0.021
Isolated Dyads	-0.948 <sup>***</sup>
Grade Dif	0.092 <sup>*</sup>
Attractiveness Difference	-0.156 <sup>**</sup>
Both Smoke	0.841 <sup>***</sup>
Popularity Difference	-0.076 <sup>*</sup>

\* p<.05, \*\*p<.001, \*\*\* pp<.0001

Appendix: Description of Individual-level Measures

Variable	Description	Range
Family SES	Occupational prestige and educational background.	0 - 10
GPA	Grade Point Average. Grades in English, Math, History and Science.	4=A 1=D or F
Expect College Graduation	How likely that you will go to college.	1=Unlikely 5=Likely
School Attachment	Mean of three items: (1) Do you feel close to people at your school, (2) are you happy to be at your school, (3) do you feel a part of your school.	1=Strongly Disagree, 5= Strongly Agree
Trouble in School	Mean of four items about relations with teachers, students and schoolwork.	0=Never 5=Everyday
Drunk	Frequency of getting drunk.	0=Never 6=Almost Every Day
Delinquency	Delinquency Scale. Mean of 15 items about theft, property damage and violence.	0=Never 5=3 or more times
Hours Watch TV	“How many hours a week do you watch TV?”	1-99 hours
Religiosity (Praying)	“How often do you pray?”	1=Never, 5=once a day.
In-Degree*	Number of times respondent is nominated as a friend.	0-37
Self Esteem	Self esteem scale. Mean of 10 items.	0=low; 4=high
Sexually Active	“Have you ever had sexual Intercourse?”	0=No; 1=Yes
Autonomy	Autonomy Scale. Sum of 7 items about Respondent’s ability to make own decisions.	0=Low Autonomy 7=High Autonomy
Expect to get AIDS	“What do you think are the chances you will get HIV or AIDS?”	1=No Chance 5=It will Happen.
Marry by 25	“What do you think are the chances you will be married by age 25?”	1=No Chance 5=It will Happen.
Attractiveness	Interviewer Assessment of Student Attractiveness.	1=Unattractive 8=Very Attractive
AH_PVT	Add Health Picture Vocabulary Test.	13 - 146
Two Biological Parents	Respondent lives with two biological parents.	0=No; 1=Yes
Smokes Regularly	Respondent smokes at least 1 cigarette a day for last 30 days.	0=No; 1=Yes
School Suspension	“Have you ever received an out-of-school suspension.”	0=No; 1=Yes
Tattoo	“Do you have a permanent tattoo?”	0=No; 1=Yes
Grade	Grade in school.	7 to 12 <sup>th</sup> grade
Major Religion.	Respondent’s religion	None, Protestant, Catholic, other.

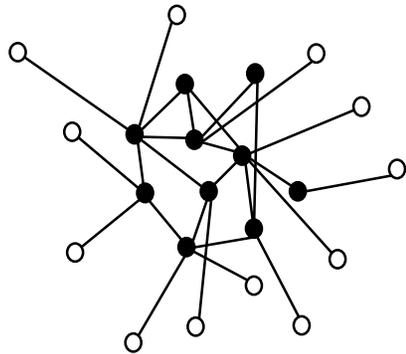
\* Measurement of *in-degree* is derived from the In-school instrument

Table 4. Robustness of the observed statistics to random rewiring

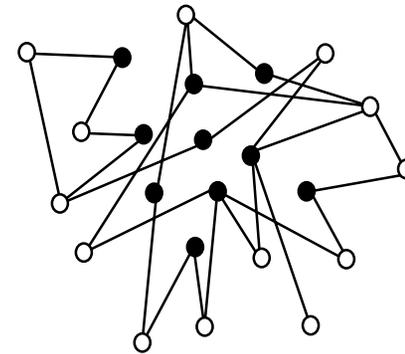
Measure	Observed	Simulated with Error				
		Mean	Min	Max	Std Dev	IQ Range
Size of largest component	288	289.83	108	344	28.5	30
Number of Components	101	103.7	96	113	2.31	3
Reachability	.254	.262	.073	.36	.042	.05
Total Avg. (Normalized) Closeness	.017	.012	.01	.029	.003	.004
Closeness (normalized) in Largest Component	.064	.072	.045	.123	.008	.011
Diameter	37	36.50	18	63	6.04	8
# of cycles	5*	8.24	1	17	2.24	3
Mean Information Centrality	.003	.004	.002	.009	.0005	.0004
Std. Dev of Information Centrality	.0007	.00076	.00047	.0019	.00012	.00012

\*Probability of observing 5 cycles given the rewiring is .07.

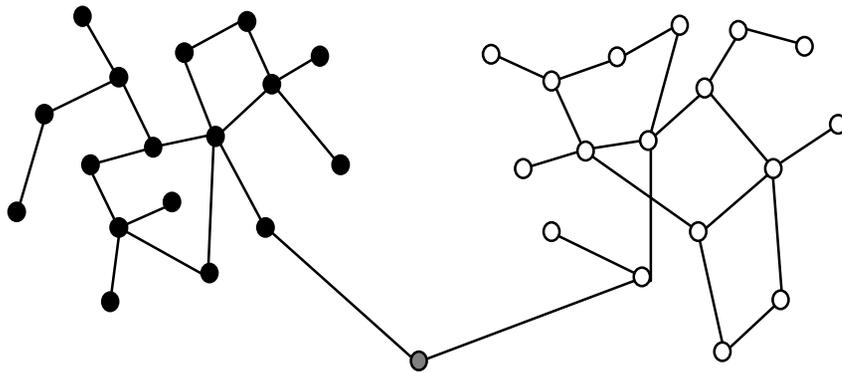
# Figure 1. Models of Contact



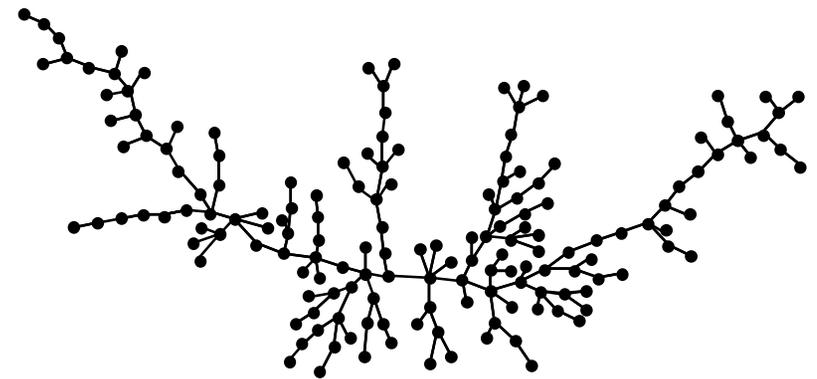
Panel A: Core Infection Model



Panel C: Inverse Core Model



Panel B: Bridge Between Disjoint Populations



Panel D: Spanning Tree

Figure 2. Structure of Romantic and Sexual Contact at Jefferson

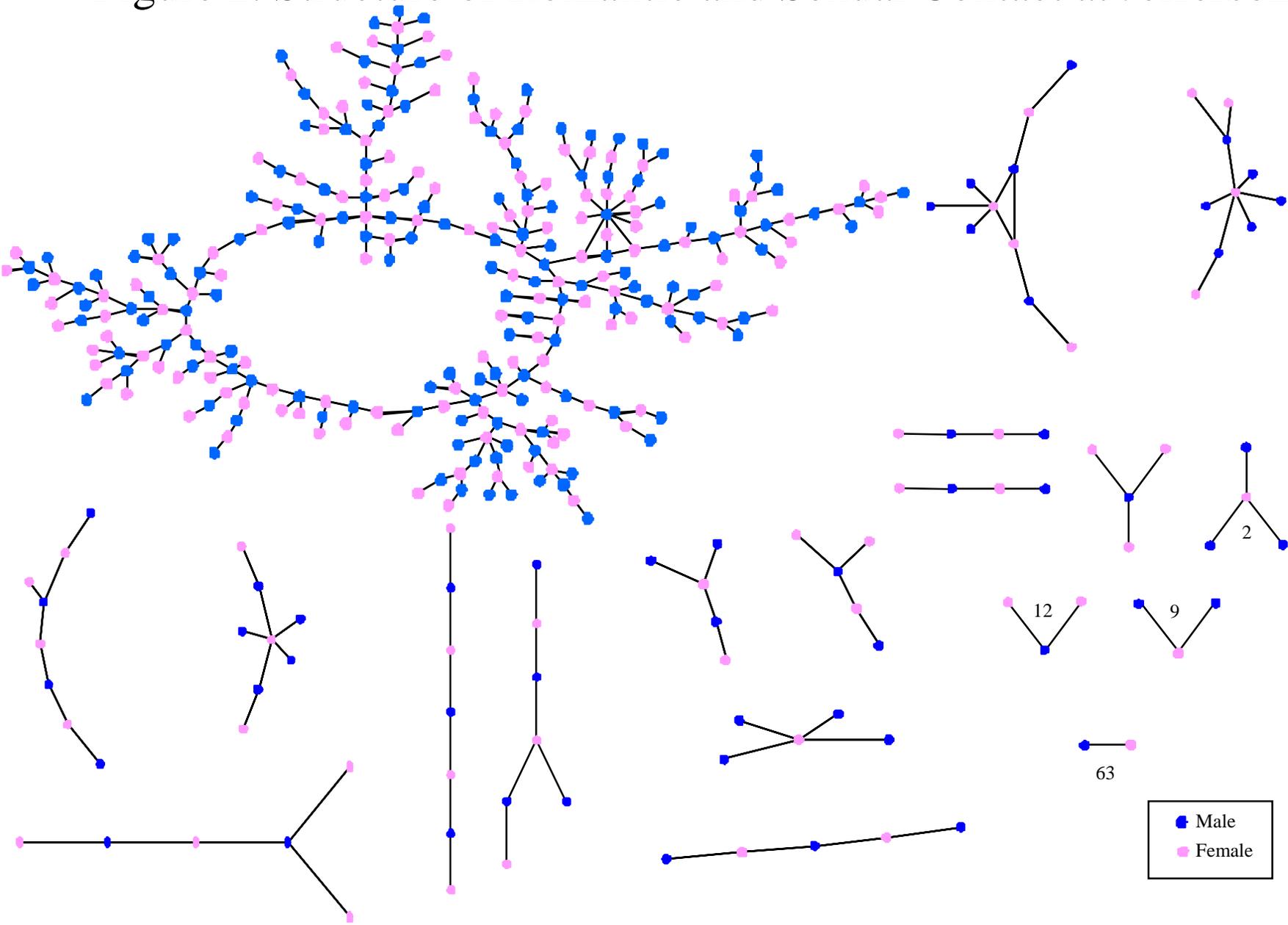


Figure 3. Indirect Pathways

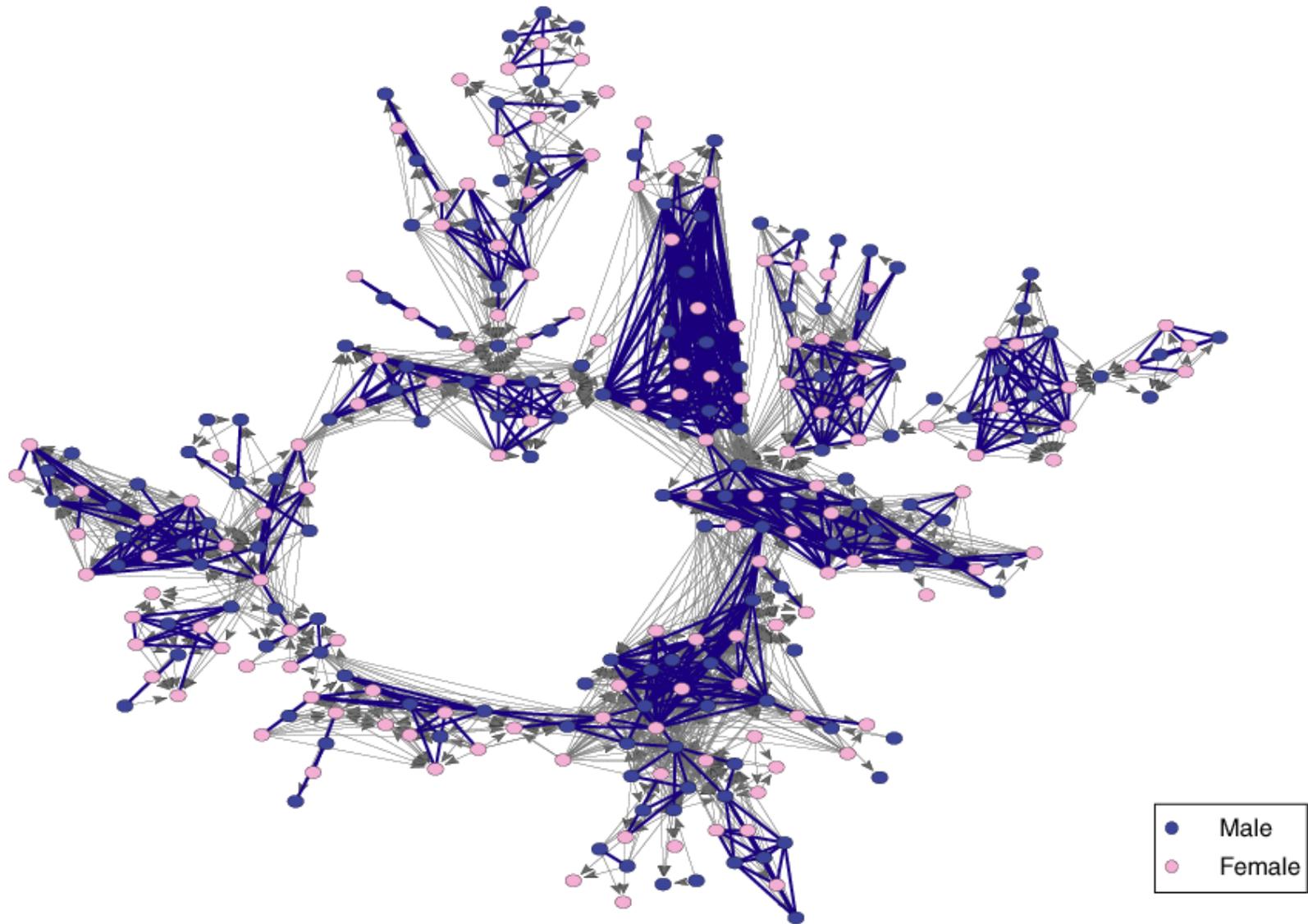
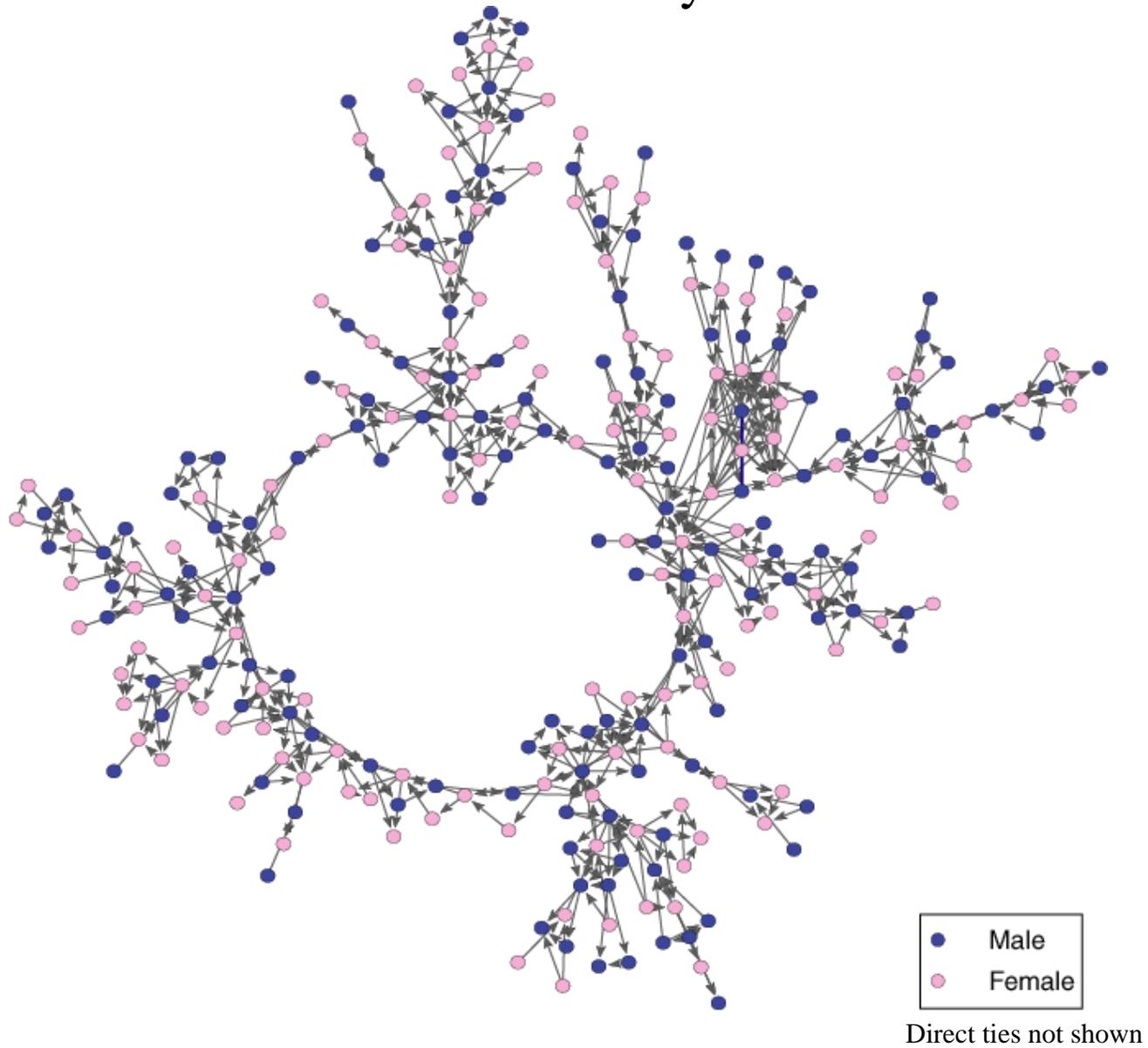


Figure 4. Minimum Indirect Pathways



# Figure 5. Simulated and Observed Network Features

Simulated Networks preserve observed degree distribution

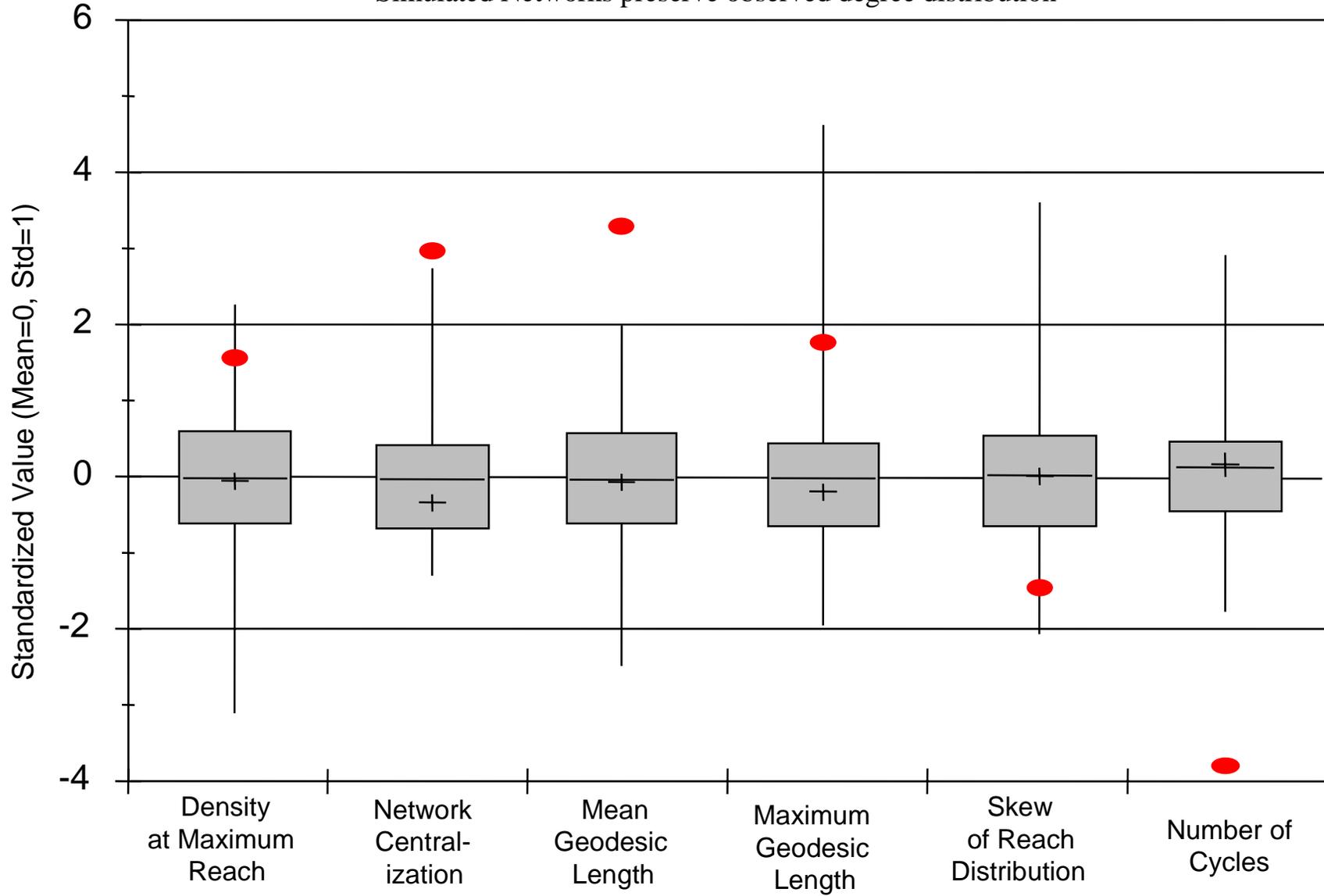
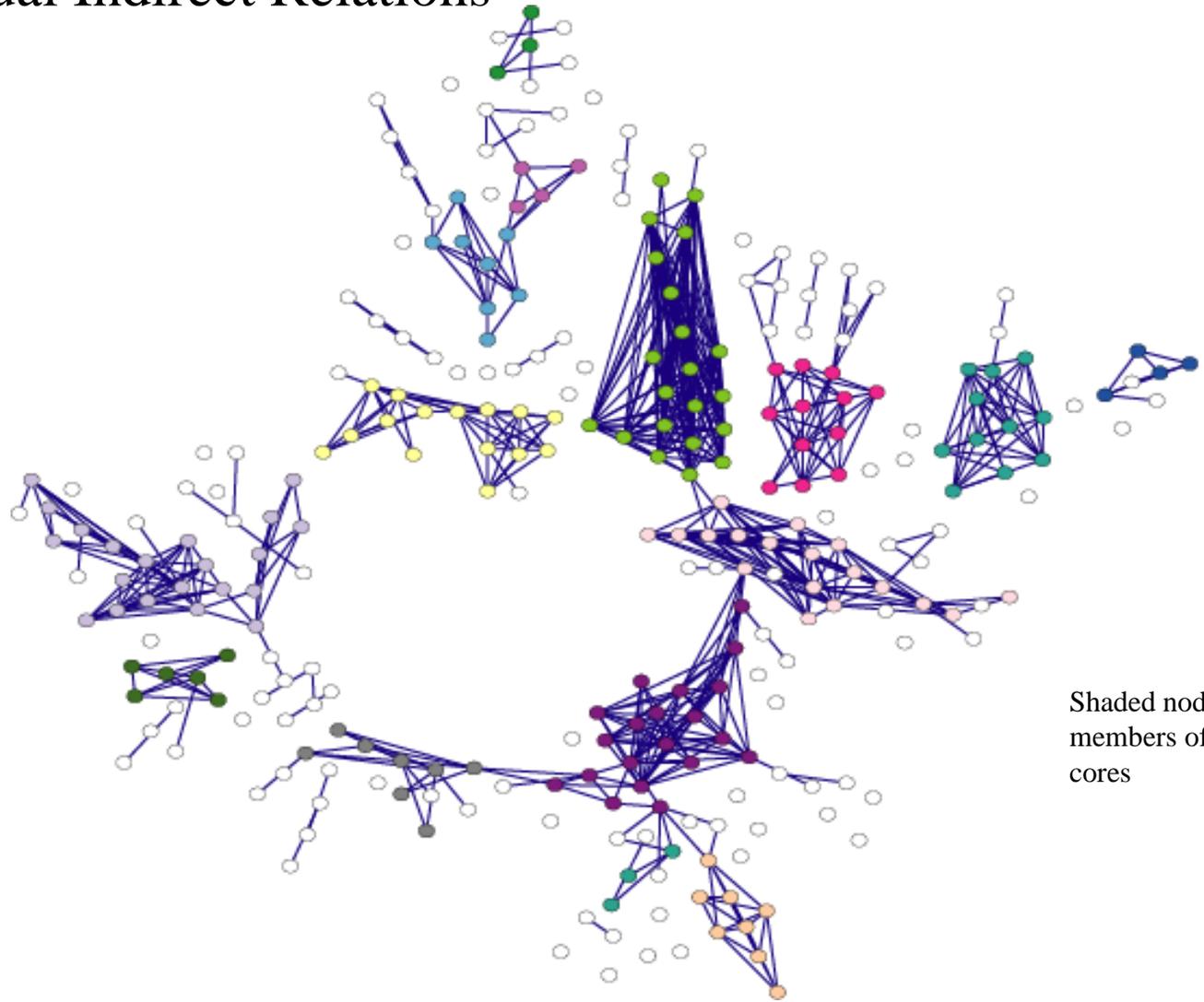


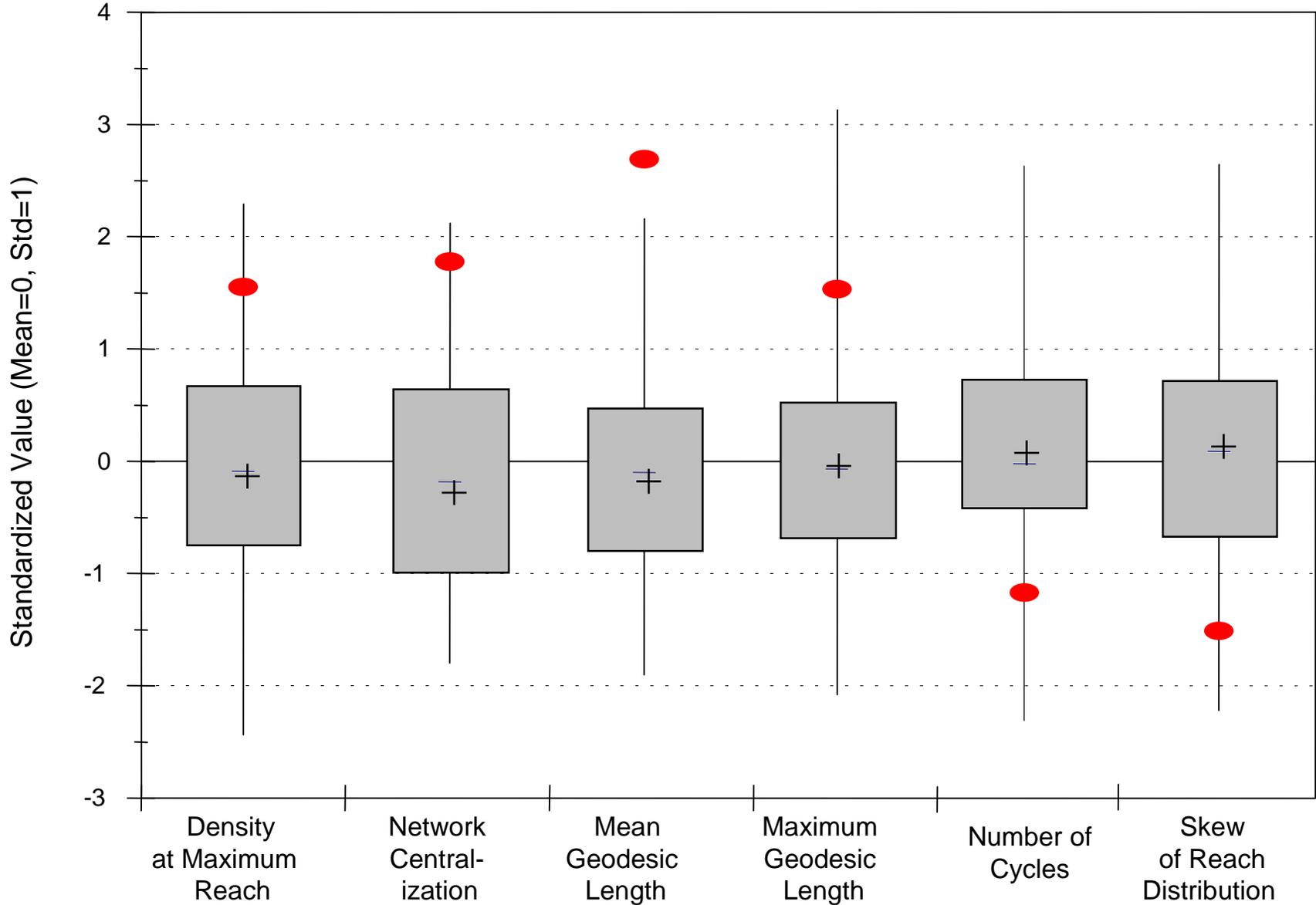
Figure 6. Indirect Cores: Bicomponents on Mutual Indirect Relations



Shaded nodes are members of indirect cores

# Figure 7. Simulated and Observed Network Features

Simulated Networks preserve observed degree and isolated dyad distribution



# Figure 8. Simulated and Observed Network Features

Simulated Networks randomly re-wire 5% of relations

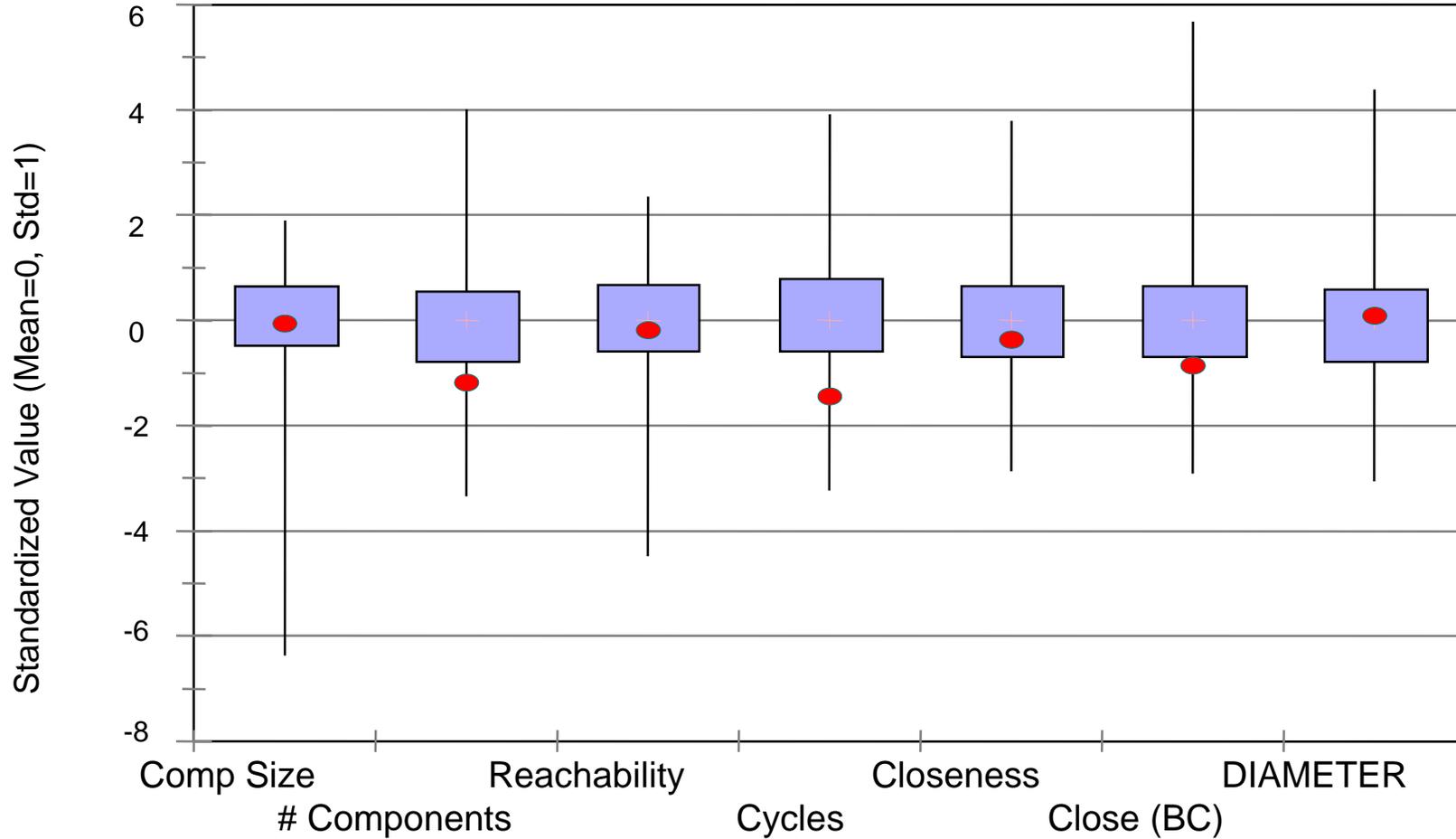
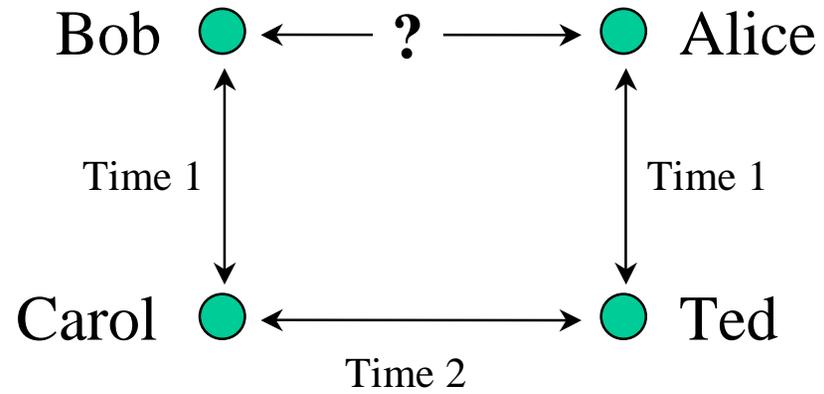
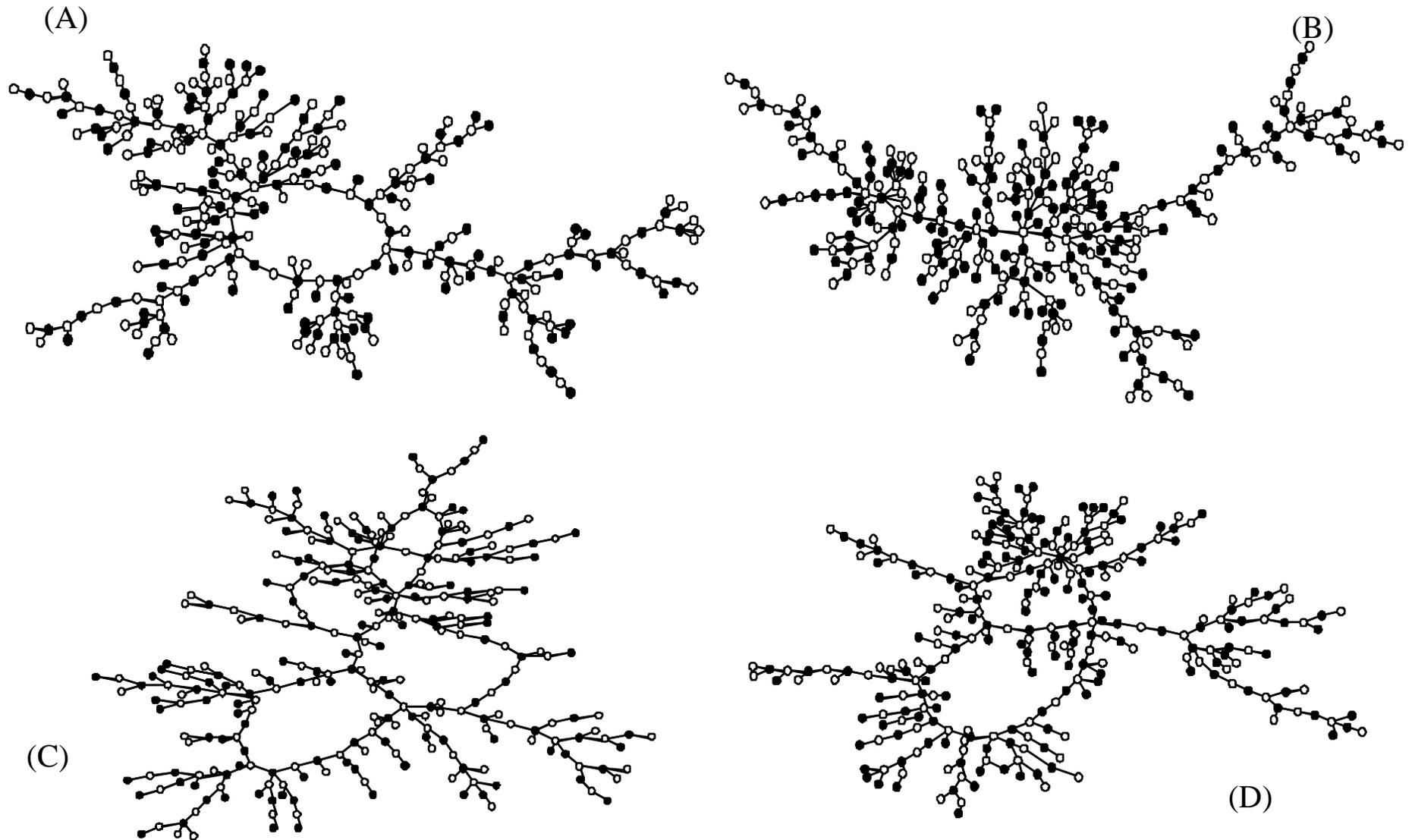


Figure 9. Hypothetical Romantic Cycle



# Figure 10. Largest Components for 4 Simulated Networks

Simulated Networks preserve observed degree, isolated dyad distribution, and cycle constraint



## **Recent ISERP Working Papers**

02-01 "Alternative Models of Dynamics in Binary Time-Series-Cross-Section Models: The Example of State Failure," Nathaniel Beck, Political Science, UC San Diego, David Epstein, Political Science, Columbia, Simon Jackman, Political Science, Stanford and Sharyn O'Halloran, Political Science, Columbia

02-02 "Substitutability Cross-Stream Between Oriented Markets: Conventions in the Wine Sector of France," Harrison White, Sociology, Columbia University

02-03 "Link, Search, Interact: The Co-Evolution of NGOs and Interactive Technology," Jonathan Bach, Center on Organizational Innovation, Columbia University and David Stark, Center on Organizational Innovation, Columbia University

02-04 "Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks," Peter Bearman, Institute for Social and Economic Research and Policy, Columbia, James Moody, Sociology, Ohio State, Katherine Stovel, Sociology, University of Washington

01-01 "Pathways of Property Transformation: Enterprise Network Careers in Hungary, 1988-2000 Outline of an Analytic Strategy," David Stark, Sociology, Columbia and Balázs Vedres, Sociology, Columbia

01-02 "Policy Space and Voting Coalitions in Congress: the Bearing of Policy on Politics, 1930-1954," Ira Katznelson, John Lapinski, and Rose Razaghian, Political Science, Columbia

01-03 "Doing Fractions: An Analysis of Partisan ship in Post-Socialist Russia," Andrew D. Buck, Sociology, Columbia

01-04 "Opposite-Sex Twins and Adolescent Same-Sex Attraction," Peter Bearman, Sociology/ISERP and Hannah Brückner, Sociology, Yale

01-05 "On the Uneven Evolution of Human Know-How," Richard R. Nelson, Business/SIPA, Columbia

01-06 "Self-Control for the Righteous: Toward a Theory of Luxury Pre-Commitment," Ran Kivetz, Business, Columbia and Itamar Simonson, Business, Stanford

01-07 "Distributing Intelligence and Organizing Diversity in New Media Projects," Monique Girard, ISERP, Columbia and David Stark, Sociology, Columbia

01-08 "Agricultural Biotechnology's Complementary Intellectual Assets," Gregory D. Graff, Agricultural and Resource Economics, Berkeley, Gordon C. Rausser, Agricultural Economics, Berkeley and Arthur A. Small, SIPA/Earth Institute, Columbia

For copies of ISERP Working Papers  
visit [http://www.iserp.columbia.edu/initiatives/working\\_papers/paper\\_program.html](http://www.iserp.columbia.edu/initiatives/working_papers/paper_program.html),  
write to [iserp@columbia.edu](mailto:iserp@columbia.edu) or call 212-854-3081.

EDITORIAL BOARD

Karen Barkey, Sociology

Peter Bearman, Sociology/ISERP

Alan Brinkley, History

Charles Cameron, Political Science

Alessandra Casella, Economics

Ester Fuchs, Political Science/SIPA

John Huber, Political Science

Ira Katznelson, Political  
Science/History

Herbert Klein, History

Mary Clare Lennon, Public Health

Mahmood Mamdani, Anthropology

Marianthi Markatou, Statistics

William McAllister, ISERP

Kathryn Neckerman, ISERP

Richard Nelson, Business/SIPA

Elliot Sclar, Architecture, Planning  
and Preservation/SIPA

Seymour Spilerman, Sociology

Charles Tilly, Sociology

Harrison White, Sociology

ADMINISTRATION

Peter Bearman, Director

Kathryn Neckerman, Associate Director

Leslie Wright, Assistant Director

Institute for Social and Economic

Research and Policy

Columbia University

International Affairs Building

420 West 118 Street, 8<sup>th</sup> Floor

Mail Code 3355

New York, NY 10027

telephone: 212-854-3081

facsimile: 212-854-8925

e-mail: [iserp@columbia.edu](mailto:iserp@columbia.edu)

URL: <http://www.iserp.columbia.edu>