Published in final edited form as:

Addict Res Theory. 2017 ; 25(3): 243–250. doi:10.1080/16066359.2016.1249864.

A network of helping: Generalized reciprocity and cooperative behavior in response to peer and staff affirmations and corrections among therapeutic community residents

Nathan J. Doogan and

The Ohio State University College of Public Health

Keith Warren

The Ohio State University College of Social Work

Abstract

Background—Clinical theory in therapeutic communities (TCs) for substance abuse treatment emphasizes the importance of peer interactions in bringing about change. This implies that residents will respond in a more prosocial manner to peer versus staff intervention and that residents will interact in such a way as to maintain cooperation.

Method—The data consist of electronic records of peer and staff affirmations and corrections at four corrections-based therapeutic community units. We treat the data as a directed social network of affirmations. We sampled 100 resident days from each unit ($n = 400$) and used a generalized linear mixed effects network time series model to analyze the predictors of sending and receiving affirmations and corrections. The model allowed us to control for characteristics of individuals as well as network-related dependencies.

Results—Residents show generalized reciprocity following peer affirmations, but not following staff affirmations. Residents did not respond to peer corrections by increasing affirmations, but responded to staff corrections by decreasing affirmations. Residents directly reciprocated peer affirmations. Residents were more likely to affirm a peer whom they had recently corrected. Residents were homophilous with respect to race, age and program entry time.

Conclusion—This analysis demonstrates that TC residents react more prosocially to behavioral intervention by peers than by staff. Further, the community exhibits generalized and direct reciprocity, mechanisms known to foster cooperation in groups. Multiple forms of homophily influence resident interactions. These findings validate TC clinical theory while suggesting paths to improved outcomes.

Keywords

Substance abuse treatment; mutual aid; therapeutic communities; evolutionary psychology; social network analysis; self-organization

Introduction

Therapeutic communities (TCs) for individuals with substance abuse and criminal behavior is based on mutual aid; the community of residents is itself the method of treatment (De Leon, 2000; Tiburcio & Kressel, 2011; Vandeveld, Broekaert, Yates & Kooyman, 2004). While all TCs include professional staff, residents work on tasks necessary for running the facility, monitor peers and provide both positive and negative feedback through a system of affirmations and corrections (De Leon, 2000; Perfas, 2012). This clinical approach is consistent with recent work that sees recovery from substance abuse as consistent with a change in identity that is facilitated by social networks and community interactions (Best, Beckwith, Haslam, Haslam, Jetten, Mawson & Lubman, 2016; Keleman, Erdos & Madacsy, 2007).

Two assumptions underlie this clinical method. The first is that residents will react to peers in a more prosocial manner than to staff (Bell, 1994; De Leon, 2000; Stevens, 2013). The second is that TC residents interact in such a way as to maintain a cooperative and safe environment (Debaere, Vanheule & Inslegers, 2014; Pearce and Pickard, 2012). This is a salient issue for TCs; the heavy dependence on peer support means that trust in peers and the social climate of the unit are primary predictors of retention (Carr & Ball, 2014; Mandell, Edelen, Wenzel, Dahl & Ebener, 2008). Moreover, recent research indicates that TC groups develop different treatment climates even when individuals are randomly assigned to different groups (Kelly & Welsh, 2016).

No quantitative research has attempted to distinguish TC resident reactions to staff versus peer intervention, and none has explored mechanisms through which residents might maintain unit climate. In this study we rigorously test the former by comparing resident responses to staff versus peer intervention. We examine the latter by testing for the presence of a set of known cooperative mechanisms in TCs. To do so, we draw on a large clinical dataset of interactions between TC residents and both staff and peers. These interactions are specific to the TC clinical environment, where it is expected that both staff and peers will intervene by affirming residents for prosocial behavior and correcting behavior that contravenes TC norms (De Leon, 2000; Perfas, 2012). The records of these interventions form a directed social network; when A affirms B there is an arrow pointing from A to B, when B affirms C there is an arrow pointing from B to C, and so on. We use longitudinal social network analysis to study these networks.

Theory and cooperation in therapeutic communities

Two bodies of theory offer insights into the cooperative functioning of TCs. The first is TC clinical theory itself, as developed over the last several decades and formalized in both books and articles (De Leon, 2000; Perfas, 2012; Stevens, 2013). The second is work on cooperation and altruism that has developed in evolutionary biology, psychology, economics and political science since the 1960s (Fu, Nowak, Christakis and Fowler, 2012; Rand & Nowak, 2013). These bodies of theory frequently parallel each other and we will consider them together, drawing three focal hypotheses and a number of ancillary hypotheses.

De Leon (2000) notes that when residents receive an affirmation from a peer they often increase prosocial activity in TCs. Such generalized reciprocity, in which individuals help third parties after having received help themselves, can stabilize cooperation in groups (Herne, Lappalainen & Kestilä-Kekkonen, 2014; Isen & Levin, 1972; Molm, Collett & Schaefer, 2007; Nowak & Roch, 2007; Rutte & Taborsky, 2008; Stanca, 2009). Warren, Doogan, De Leon, Phillips, Moody & Hodge (2013) found evidence of generalized reciprocity of affirmations in TCs from one week to the next. However, the model used negative binomial regression in which the aggregated number of affirmations sent per week was the dependent variable, which made it impossible to control for network interactions between individuals, particularly direct reciprocity (affirmations sent back to a sender). Focal hypothesis one is therefore that *residents will send more affirmations after receiving a peer affirmation*, when controlling for other individual interactions.

Peer *sanctions*, including verbal corrections, can also increase cooperation (Barr, 2001; Bowles & Gintis, 2011; Fehr & Gächter, 2002; Masclet, Noussair, Tucker & Villeval, 2003). TCs make use of peer verbal corrections, which are seen as an expression of responsible concern for fellow residents who may be showing behavior that is destructive to themselves or others (De Leon, 2000; Perfas, 2012). Warren et al (2013) found that the number of peer corrections that TC residents received during one week predicted the number of affirmations they sent the next. Again, not all possible confounds were controlled because of the negative binomial model. Focal hypothesis two is that *residents will send more affirmations after receiving a peer correction*.

Cooperation between peers with more or less equal initial power in the group is of interest to evolutionary theorists because the idea that individuals will respond to a legal authority is considered unproblematic. TC clinical theory goes one step further, arguing that residents will actually respond more strongly to peer intervention than staff intervention; this is the rationale for regarding the entire community as the method of treatment (De Leon, 2000; Perfas, 2012; Stevens, 2013). Focal hypothesis three is that *residents will show a stronger increase in sending affirmations in response to peer affirmations or corrections than in response to staff affirmations or corrections*.

These hypotheses allow us to test the core TC clinical claims that residents do respond to peer intervention and that they respond preferentially to peer versus staff intervention. Statistical testing of hypotheses in a social network framework requires controlling for network structure, since the observations are not independent (Wasserman & Faust, 1994). We therefore propose a series of control variables, most of which are also of clinical interest. First, there is both theoretical and experimental evidence that reciprocity, common in human social interactions, can stabilize cooperation in large groups (Axelrod, 1984; Nowak, 2006; Falk & Fischbacher, 2006; Rand & Nowak, 2013). If left uncontrolled, reciprocity can inflate the generalized reciprocity estimates; moreover, qualitative research indicates that TC residents value reciprocity in helping (Miller, Sees & Brown, 2006; Stevens, 2013). We expect that residents will reciprocate affirmations.

Second, when peers reward those who are known to cooperate, cooperation in the group as a whole can be more stable; reputation effects may therefore play an important role in

maintaining cooperation in groups (Rand & Nowak, 2013). There is evidence that reputation effects play a role in TC treatment (Hodge, Warren & Linley, 2014; Miller et al, 2006; Stevens, 2013). We expect that residents who send more affirmations will also receive more as a reputational reward for their activity in the TC.

Third, De Leon (2000) suggests that residents attempt to balance corrections sent to peers with affirmations. This could play out in three ways. First, all residents could be more likely to affirm a peer who has recently received a correction. Second, a resident could be more likely to affirm a particular peer whom he or she has recently corrected. Third, a resident could be more likely to send an affirmation to any peer after sending a correction, so as to avoid appearing overly negative. We expect to find these three relationships.

We expect two other network relationships. First, *transitivity* is a network construct in which two individuals are more likely to interact because of a common third (Wasserman & Faust, 1994). This could confound our main hypotheses focused on generalized reciprocity. Second, uncontrolled autocorrelation can shrink the confidence interval around parameter estimates, thereby leading to Type I error. Autocorrelation is also of clinical interest because it would imply that resident activity has momentum. Autocorrelation in sending affirmations to all peers, sending affirmations to specific peers, receiving affirmations from peers and remaining an isolate is expected.

Homophily is the principle that individuals connect with those who are like themselves (Goodreau, Kitts & Morris, 2009; McPherson, Smith-Lovin & Cook, 2000; Rivera, Soderstrom & Uzzi, 2010). It is of concern in TCs because of the possible formation of trait-based cliques that could counter the clinical action of the entire community (De Leon, 2000). The analysis tests for homophily by age, race, risk of recidivism and time in program. Individual characteristics might also determine how many links one has, regardless of whom one is linked to. Age, time in program, race and risk of reincarceration are therefore included as individual level predictors of sending and receiving affirmations.

Methods

Data

All data were gathered under The Ohio State University IRB protocol 2009B0174. Data for the study were drawn from archival records kept at four male correctional TC units at three TCs in Ohio. Two of the TCs drew on predominantly rural catchment areas. Of these, one had a 90-bed male unit while the other was a 60-bed entirely male facility. The third TC drew on a mixed urban/rural catchment area. This TC included two 80-bed male units. All of the TCs were short-term, with a maximum stay of six months. All residents were felony criminal offenders who had either been diverted from a longer sentence or transferred to the TCs for treatment at the end of a prison sentence. Treatment was focused on reducing criminal recidivism and substance abuse. Resident interactions were recorded as a means of monitoring clinical operations and maintaining fidelity to the TC principle that the community should be the primary method of treatment (De Leon, 2000). Data included a total of 353,294 affirmations exchanged between 4,634 residents; this article uses a random subsample of this dataset.

Both affirmations and corrections were recorded in three steps. First, the sender wrote the names of the sender and receiver, the date, and the content on a form. Second, senior residents and staff reviewed the form. Finally, the affirmation or correction was read aloud to the community and then entered into a database. The result was an electronic record of daily affirmations and corrections. While residents were individually identified in the dataset, staff were not. The staff included a mixture of paraprofessionals who were responsible for correctional custody and professionals such as credentialed social workers, counselors and teachers.

These facilities also kept data on ethnicity, age, entry date and residents' scores on the Level of Services Inventory-Revised (LSI-R), a standardized measure of risk of criminal recidivism (Andrews & Bonta, 1995). The LSI-R includes information on substance abuse, education level, previous offenses, employment, financial status, social support, family/marital status, living accommodations, recreational skills, mental health issues and attitudes toward criminal behavior. Ethnicity was recoded into a bivariate dummy variable with minority status as the reference value.

Sampling

The unit of analysis is the directed dyad—resident to resident or staff to resident—in which the event of interest is whether or not an affirmation occurred originating from the sending individual and sent to the receiving individual. We randomly sampled 100 days from each TC unit within the time period from January 2006 to September of 2008—a period within which all four TCs were operating and collecting data. The number of days sampled and the size of the TC units meant a massive number of directed dyad days, in which one resident could affirm another. For example, one unit contained 90 beds and was represented by 100 days in the data. The total number of directed dyad days was therefore $90 \text{ senders} \times 89 \text{ receivers} \times 100 \text{ days}$, or 801,000. The total number of records needed to represent all dyads for all days from all TCs was approximately 2,064,000. But only 12,104 affirmations were actually exchanged during these days. King and Zeng (2001) recommend that at least five percent of records represent the occurrence of events to avoid biases in parameter estimation. We adopted their suggested solution, retaining all events, and randomly sampling from non-events such that 20% of all data records represented a directed dyad in which an affirmation or correction occurred.

Analysis

To capture the complex dependencies in the data, we adapt the approaches of de Nooy, (2012) and Lerner, Bussmann & Snijders (2013), building a generalized linear mixed effects model. We go into detail about the time-lagged variables that capture network dependencies in a later section.

We incorporate random effects into the model to capture within-person correlations in propensity to send affirmations/corrections. Any dyad in which an affirmation may occur is composed of both a sender and a receiver. We capture this within-person correlation in the propensity of the receiver to *receive* affirmations/corrections with a second random effect. Finally, many events may occur within a single day and the number varies. We therefore

include a third random effect to capture variation across days in the proportion of dyads in which an event occurs.

The model is fit using a Markov chain Monte Carlo (MCMC) sampler called STAN (Stan Development Team, 2015). STAN is a Bayesian modeling framework, allowing the incorporation of *a priori* knowledge (a.k.a., a “prior”) into the model parameter estimates (Gelman, Carlin, Stern & Rubin, 2004). We use zero centered and diffuse priors for all fixed mean parameters (normal distribution with standard deviation of five). If the prior has any bearing on the parameter estimates, it shrinks them towards zero. However, this shrinkage effect is likely to be infinitesimal due to the large number of observations and the weak information encoded into the prior. A uniform prior is used for the random effect variances as suggested by Gelman (2006). We characterize the posterior distribution of the parameters by the mean of each parameter and a 95 percent equal-tailed Bayesian credible interval (95% CI). We determine the statistical significance of all quantities reported by whether or not their 95% CIs cross zero.

Measures

Dependent variable—The model is of the probability of an affirmation from an actor i to actor j . The event from actor j to actor i is a distinct data point involving the same two individuals. Observations are also identified by the day on which they have occurred, t . Because a total of 400 days are sampled from four TC units, the potential event of the ij dyad may be represented in the data several times, though always on different days indexed by t .

Time decay—We assume that the influence of peer and staff affirmations and corrections decays over time. The weight w_{ijt} of a specific event—originating from actor i , destined for actor j , during time period t —is the value of the weight of the same event at the previous time period w_{ijt-1} multiplied by the constant scalar decay factor δ and then added to 1 if the event reoccurred in the previous time period and added to 0 if it did not. Mathematically,

$$w_{ijt} = \delta w_{ijt-1} + y_{ijt-1}$$

where δ is the decay factor that can take on real values in the range from 0 (decay to 0 in one time period) to 1 (no decay), and y is the dichotomous dependent variable.

Table 1 shows the formulas for each type of independent variable, some of which depend on the history of events encoded in w . In the table, w has three subscripts. A dot (.) is used to indicate a sum across an entire dimension of the w array; therefore, $w_{.ib}$ the formula for in-degree activity type effects (explained in more detail below), means the sum of the weights of ties sent by all actors to actor i at time t . In other words, this is the total in-degree of actor i at time t accounting for the decay of historical ties. The actual decay rate used in models was chosen to minimize the model deviance.

Nomenclature for time lagged network predictors—Time lagged network predictors measure the relationship between historical-network characteristics and the probability of a

new affirmation. The naming scheme is drawn from social network terminology (Wasserman & Faust, 1994). “*Indegree*” is the sum of historical affirmations/corrections received by a resident and “*outdegree*” is the sum of historical affirmations/corrections sent by a resident. Independent network variables suffixed with “*activity*” affect the probability of the dyadic event as a function of a characteristic of the sender in the dyad (the one creating the *activity*). If it represents the indegree of the sender, the variable name will be some form of the term “*indegree activity*.” Independent variables suffixed with “*popularity*” affect the probability of the event as a function of a characteristic of the receiver (who has a *popularity* in receiving affirmations or corrections). Independent variables that predict homophily are called *similarity* or *same* variables. Independent variables that adjust the probability of one type of event (e.g., affirmations) based on information about the history of a different type of event (e.g., corrections) are called *cross-network* variables or include both network types (“affirmation” and “correction”) in their name. For example, *correction-indegree affirmation-activity* refers to a cross-network variable that adjusts the probability of an affirmation sent to anyone as a function of the correction indegree of the sender.

Generalized reciprocity and response to peer correction—All of the generalized reciprocity and response to correction parameters are *indegree activity* type variables. The model will include four variables that capture the adjustment in the probability of giving an affirmation after receiving an affirmation from a peer (*affirmation-indegree affirmation-activity*), a correction from a peer (*correction-indegree affirmation-activity*), receiving an affirmation from staff (*staff affirmation-indegree affirmation-activity*) or a correction from staff (*staff correction-indegree affirmation-activity*).

Tests compare the parameter estimates representing the response to peer versus staff affirmation exposure and peer versus staff correction exposure. For example, one test compares the parameter estimate for *affirmation-indegree affirmation-activity* to the parameter estimate for *staff affirmation-indegree affirmation-activity*.

Control of alternative explanations

Reciprocity—If resident A affirms or corrects resident B, it is expected that resident B will be more likely to affirm or correct resident A. The model therefore includes a *reciprocity* variable and a cross-network form, *cross-network reciprocity*, capturing reciprocal affirmations in response to correction.

Reputation for affirming peers—The reputation variables are called *X-outdegree Y-popularity* (e.g., *affirmation-outdegree affirmation-popularity*) and we include a cross-network form (e.g., *correction-outdegree affirmation-popularity*).

Affirmation/correction sending balance—TC residents might be more likely to send an affirmation after sending a correction in order to manage their reputation and avoid seeming too negative. The variable capturing this activity balance is *correction-outdegree affirmation-activity*.

Affirmation/correction receiving balance—If resident A corrects resident B, others may be more likely to affirm resident B to *communally* balance B’s experience (*correction-*

indegree affirmation-popularity and *affirmation-indegree correction-popularity*). The analysis also controlled for the likelihood that a resident would affirm a *specific* peer whom he had previously corrected, a focused version of the balance hypothesis. These variables are called *cross-network concentration* (“concentration” being a focus on one individual).

Network structural controls—Social network structural controls were included to account for confounding relationships (Pedhazur & Schmelkin, 1991). In addition to *transitivity*, four types of autocorrelation are controlled: If A sends an affirmation today he is more likely to send one tomorrow, (*X-outdegree X-activity*). If A receives an affirmation or correction today, he is more likely to receive one tomorrow, (*X-indegree X-popularity*). If A affirms B today he is more likely to affirm B tomorrow (*X-concentration*). If A has sent no recent affirmations, he will continue the trend (*X-isolation X-activity*). (X is a placeholder for “affirmation” or “correction.”)

Individual characteristics—Four controls for homophily were included: race (*same race*), age (*age similarity*), LSI-R score (*LSI-R similarity*), and days since admission (*time similarity*). The analysis controlled for the same four individual attributes that could influence the likelihood of an affirmation occurring using both *covariate activity* and *covariate popularity* type variables. For these, age was continuous and race was dichotomous (minority = 0; European American=1). Time in the program was recoded into two dummy variables indicating whether a resident was in his first month of the program or past his fourth month.

Results

Descriptive statistics are found in Table 2. Seventy-two percent of residents were European American, with a mean age of 28.78 and a mean LSI-R score of 28.14. Mean length of stay was 140.13 days, with a minimum length of stay of zero (the resident had arrived that day) and a maximum length of stay of 184 (one day over the six-month program maximum). Descriptive statistics indicate that residents receive far more affirmations and corrections on average from their peers than from staff.

Model

The results of the model are found in Table 3. Residents who had received an affirmation from a peer were more likely to affirm a third peer (*affirmation-indegree affirmation-activity*: $b = .177$, 95%CI [.123, .228]). In contrast, residents were not significantly more or less likely to send an affirmation after receiving a correction from a peer (*correction-indegree affirmation-outdegree*: $b = .045$, 95%CI [−.012, .103]).

The *staff affirmation-indegree affirmation-activity* parameter was not statistically significant ($b = .035$, 95%CI [−.028, .104]), yielding no evidence that residents were more likely to affirm peers after receiving an affirmation from staff. Response to peer affirmations is significantly more positive than the response to those sent by staff (difference = .142, 95%CI [.053, .234])

Staff correction-indegree affirmation-activity was negative and statistically significant ($b = -.125$, 95%CI $[-.197, -.050]$); residents were less likely to affirm a peer after receiving a correction from a staff member. The response to peer corrections (which was not statistically significant) is significantly more positive than the negative response to staff corrections (difference = $.170$, 95%CI $[.076, .259]$).

Reciprocity

Residents reciprocated affirmations (*Reciprocity*: $b = .774$, 95%CI $[.614, .941]$), but were no more likely to affirm a peer who corrected them (*Cross-network reciprocity*: ($b = -.073$, 95%CI $[-.347, .209]$).

Engagement-based popularity

Residents were no more likely to receive an affirmation after sending an affirmation (*Affirmation-Outdegree Affirmation Popularity*: $b = -.004$, 95%CI $[-.041, .033]$) or after sending a correction (*Correction-Outdegree Affirmation Popularity*: $b = .028$, 95%CI $[.087, .181]$).

Affirmation and correction sending/receiving balance

Residents showed an increased tendency to send affirmations after having sent corrections (*Correction-Outdegree Affirmation-Activity*: $b = .136$, 95%CI $[.087, .181]$). Residents were more likely to affirm a specific peer whom they had recently corrected (*Cross-Network Concentration*: $b = .310$, 95%CI $[.088, .538]$). Residents were not more likely to affirm a peer who had received a correction from someone else (*Correction-Indegree Affirmation-Popularity*: $b = -.014$, 95%CI $[-.075, .044]$).

Network structural controls

Results showed no evidence of transitivity (*Transitivity*: $b = -.008$, 95%CI $[-.114, .093]$). All controls for autocorrelation were statistically significant. Residents who had recently sent an affirmation were more likely to send another (*Affirmation-outdegree affirmation-activity*: $b = .409$, 95%CI $[.374, .445]$). Residents who had already affirmed a specific peer were more likely to affirm that peer again (*Concentration activity*: $b = .237$, 95%CI $[.099, .372]$). Residents who had recently received an affirmation were more likely to receive another (*Affirmation-indegree affirmation-popularity*: $b = .125$, 95%CI $[.066, .189]$). Isolation with respect to affirmation giving resulted in a lower likelihood of sending an affirmation (*Affirmation-Isolation Affirmation-Activity*: $b = -.626$, 95%CI $[-.723, -.534]$).

Individual characteristics

The analysis found three sources of homophily in affirmations: *Time Similarity*, in which residents were more likely to affirm peers who had been in the program for similar lengths of time, ($b = 2.569$, 95%CI $[2.178, 2.937]$), *Age Similarity* ($b = .331$, 95%CI $[.149, .520]$) and *Same Race* ($b = .333$, 95%CI $[.265, .406]$).

Three individual characteristics played a role in residents' *activity* or *popularity*. European-American residents were less likely to receive affirmations than minority residents (*EuroAmerican Popularity*: $b = -.119$, 95%CI $[-.216, -.028]$), but not less likely to send

them (*EuroAmerican Activity*: $b = -.083$, 95%CI $[-.220, .058]$). Residents with higher *LSI-R* scores were less likely to send an affirmation (*LSI-R Activity*: $b = -.015$, 95%CI $[-.024, -.006]$). Residents were less likely to send affirmations in their first month in the program and after their fourth month in the program (*First Month Activity*: $b = -.192$, 95%CI $[-.274, -.106]$; *Greater than 4th Month Activity*: $b = -.337$, 95%CI $[-.418, -.259]$). They were not significantly more or less likely to receive affirmations in the first month of their program (*First Month Popularity*: $b = .044$, 95%CI $[-.026, .114]$) but were less likely to receive affirmations from the fourth month on (*Greater than 4th Month Popularity*: $b = -.208$, 95%CI $[-.276, -.139]$). Age did not alter the odds of a resident either sending or receiving an affirmation at a statistically significant level.

Discussion

As predicted, residents increased odds of sending an affirmation after receiving a peer affirmation. There was no evidence that they did so after receiving a staff affirmation. Residents did not increase the odds of sending an affirmation following peer corrections, but the odds decreased following staff corrections. In both cases, therefore, residents were more likely to affirm a peer following peer intervention than following staff intervention, and the difference was statistically significant. Direct reciprocity with peers was also evident, and residents were more likely to affirm a peer whom they had recently corrected. There was evidence of autocorrelation in both sending and receiving affirmations; however, residents who sent an affirmation to a peer were no more likely to receive affirmations apart from the direct reciprocity mentioned above. Residents who were isolates were likely to remain so. The analysis found homophily by age, race and entry time. Residents sent more affirmations in the middle of their stay than at the beginning or the end. European-Americans received fewer affirmations than minority residents.

The findings from this analysis have rich implications for TC clinical theory and practice. There is clear evidence of generalized reciprocity of peer affirmations, but not of affirmations received from staff. This is consistent with De Leon's (2000) assertion that receiving affirmations can increase program activity. It suggests that TC residents will more actively engage in treatment if clinicians or staff intervene through mediating community members rather than directly. Further, there is evidence that, while receiving help may reduce self-competence, generalized reciprocity increases it (Alvarez & van Leeuwen, 2015), and that self-competence correlates with hopefulness (Ferrari, Stevens, Legler & Jason, 2012).

While this analysis does not directly test the effectiveness of corrections in reducing problem behaviors, it does indicate that staff corrections have the unwanted side effect of reducing at least one prosocial behavior, affirmations, while *peer* corrections do not have the same effect. This again supports a clinical emphasis on the community as the method of treatment (De Leon, 2000). When staff do correct residents they should be aware that a subsequent reduction in resident activity is possible; if this leads to further corrections it could establish a feedback loop that is detrimental to resident participation.

The analysis demonstrated a strong effect size for direct reciprocity, a mechanism known to foster cooperation in groups (Rand & Nowak, 2013) and previously noted in TC clinical theory and qualitative research (De Leon, 2000; Loat, 2006; Miller, Sees & Brown, 2006; Stevens, 2013). The presence of reciprocity in affirmations suggests that the reciprocal support that underlies much of TC treatment may be relatively easy to build. In fact, it is likely an extension of a deep-seated human tendency to reciprocate help (Axelrod, 1984; Rand & Nowak, 2013). One proviso should be noted. While TC clinical theory views reciprocity in a positive light, Salter (2004) has argued that criminal offenders both observe norms of direct reciprocity and use them to manipulate others. It is therefore unclear when reciprocity is desirable.

Further, residents were more likely to affirm peers whom they had previously corrected. This confirms De Leon's (2000) hypothesis that TC residents use affirmations to balance corrections and suggests that TC residents attempt to maintain relationships with specific peers by affirming those they have personally corrected. Residents were also more likely to affirm peers whom they had not corrected after sending a correction; they may have been trying to manage their reputations so as to not appear too negative.

Clinically, autocorrelation in sending and receiving affirmations will show up as streakiness in residents' program activity. It is of particular concern that isolates were less likely to affirm peers, even when controlling for individual characteristics. Thus, those who were not connected to peers did not take action to change their status as isolates. Consistent with the qualitative findings of Ronel, Chen & Elisha (2015), this subgroup of TC residents may require additional intervention to support connection with the community.

The analysis clearly demonstrates that homophily plays a role in the way in which TC residents interact. While racial homophily in affirmations will come as no surprise—it is ubiquitous in American society and American correctional institutions (Phillip, 2008; Trammel, 2009)—the analysis revealed a more complex pattern. Residents were more likely to affirm peers of similar age and peers who had entered the TC at about the same time. This analysis suggests that three likely axes along which cliques (De Leon, 2000) may form are race, age and time of entry into the TC.

Cohort homophily—in the tendency for residents who entered the program at a similar time to interact—is of particular interest. It is easy to see that residents who arrive at about the same time could develop trust through shared experiences in the TC (Stevens, 2013). This suggests that reciprocity, discussed above, may be a sign of mutual trust rather than manipulation (Salter, 2004; Stevens, 2013). It is also possible that TC clinicians could use cohort homophily to counter racial homophily by consciously admitting racially mixed groups.

The analysis also found that senior residents were less likely to affirm peers than residents who were mid-way through their stays. Senior residents in TCs may require encouragement to stay engaged, particularly with more junior residents. On the other hand, the finding that the most junior residents were less active in affirming peers is consistent with TC clinical theory, which emphasizes that residents should begin by learning from peers (De Leon,

2000; Stevens, 2013). European-American residents were slightly less likely to receive affirmations from peers, as were residents at a higher risk of reincarceration. While neither group sent fewer affirmations, the fact that European-Americans received fewer may indicate less engagement in other aspects of the TC.

Limitations

Several limitations should be noted. While longitudinal, the data are observational in nature, which limits the causal conclusions that can be drawn. But the hypotheses themselves are drawn from principles well established in behavioral economics, evolutionary biology and sociology, while the network structure of the hypotheses means that they can be stated with unusual specificity. This strengthens the internal validity of the analysis (Hill, 1964). There are limitations to external validity stemming from the location of the facilities and the entirely male sample. Because the data set did not identify individual staff members, we cannot rule out the possibility that the relationship of staff affirmations to resident affirmations might depend in part on the training of the staff.

Concluding remarks

Many questions remain to be answered. For instance, do residents who show a greater propensity for generalized reciprocity have better outcomes? Do different patterns of response to peers lead to different positions in the social network of all residents, and do these positions predict outcomes? Longitudinal social network analysis can answer these questions and others; it gives us a quantitative tool to analyze TCs as they are, complex and constantly changing communities of residents who help each other in recovery.

Acknowledgments

The authors would like to acknowledge support from NIH grant R21DA023474, and valuable conversations with George De Leon, Ian Hamilton and Jessica Linley.

References

- Alvarez K, Leeuwen E. Paying it forward: How helping others can reduce the psychological threat of receiving help. *Journal of Applied Social Psychology*. 2015; 45(1):1–9.
- Andrews, DA., Bonta, J. The level of service inventory-revised. Toronto, Canada: Multi-health systems incorporated; 1995.
- Andrews, DA., Bonta, J. The psychology of criminal conduct. 5. New Providence, NJ: Matthew Bender & Co.; 2010.
- Axelrod, RM. The evolution of cooperation. New York: Basic Books; 1984.
- Barr, A. Social dilemmas and shame-based sanctions: Experimental results from rural Zimbabwe. Retrieved from Oxford: 2001.
- Bell DC. Connection in therapeutic communities. *International Journal of the Addictions*. 1994; 29(4): 525–543. [PubMed: 8188445]
- Best D, Beckwith M, Haslam C, Haslam HA, Jetten J, Mawson E, Lublum DI. Overcoming alcohol and other drug addiction as a process of social identity transition: the social identity model of recovery (SIMOR). *Addiction Research and Theory*. 2016; 24(2):111–123.
- Bowles, S., Gintis, H. A cooperative species: human reciprocity and its evolution. Princeton: Princeton University Press; 2011.

- Carr WA, Ball SA. Predictors and treatment outcomes of perceived ward atmosphere among therapeutic community residents. *Journal of Substance Abuse Treatment*. 2014; 46(5):567–573. [PubMed: 24560437]
- Debaere V, Vanheule S, Inslegers R. Beyond the ‘black box’ of the therapeutic community for substance abusers: A participant observation study on the treatment process. *Addiction Research & Theory*. 2014; 22(3):251–262.
- De Leon, G. *The therapeutic community: theory, model, and method*. New York: Springer Pub; 2000.
- De Nooy W. Networks of action and events over time. A multilevel discrete-time event history model for longitudinal network data. *Social Networks*. 2011; 33(1):31–40.
- Falk A, Fischbacher U. A theory of reciprocity. *Games and Economic Behavior*. 2006; 54(2):293–315.
- Fehr E, Gächter S. Altruistic punishment in humans. *Nature*. 2002; 415(6868):137–140. [PubMed: 11805825]
- Ferrari JR, Stevens EB, Legler R, Jason LA. Hope, self-esteem and self-regulation: Positive characteristics among men and women in recovery. *Journal of Community Psychology*. 2012; 40(3):292–300. [PubMed: 25382880]
- Fu F, Nowak MA, Christakis NA, Fowler JH. The evolution of homophily. *Nature Scientific Reports*. 2012; 2 [September 15, 2015] article 845, downloaded from <http://www.nature.com/articles/srep00845-message-global=remove>.
- Gelman, A., Carlin, JB., Stern, HS., Rubin, DB. *Bayesian data analysis*. Boca Raton, Fla: Chapman & Hall/CRC; 2004.
- Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*. 2006; 1(3):515–534.
- Goodreau SM, Kitts JA, Morris M. Birds of a Feather, or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks. *Demography*. 2009; 46(1): 103–125. [PubMed: 19348111]
- Herne K, Lappalainen O, Kestilä-Kekkonen E. Experimental comparison of direct, general, and indirect reciprocity. *The Journal of Socio-Economics*. 2013; 45:38–46.
- Hodge A, Warren K, Linley J. Predictors of resident role model status in therapeutic communities. *Therapeutic Communities: The International Journal of Therapeutic Communities*. 2014; 35(4): 159–167.
- Isen AM, Levin PF. Effect of feeling good on helping: Cookies and kindness. *Journal of Personality and Social Psychology*. 1972; 21(3):384–388. [PubMed: 5060754]
- Keleman G, Erdos MB, Madacsy J. Voices of sobriety: Exploring the process of recovery through patient testimonials. *Addiction Research and Theory*. 2007; 15(2):127–140.
- Kelly CE, Welsh WN. Examining treatment climate across prison-based substance abuse treatment groups. *Substance Use & Misuse*. 2016; 51(7):902–911. [PubMed: 27144315]
- King G, Zeng L. *Explaining Rare Events in International Relations*. International Organization. 2001; 55(03):693–715.
- Lerner J, Bussmann M, Snijders TAB. Modeling frequency and type of interaction in event networks. *Corvinus Journal of Sociology & Social Policy*. 2013; 4(1):3–32.
- Mandell W, Edelen MO, Wenzel SL, Dahl J, Ebener P. Do dimensions of therapeutic community treatment predict retention and outcomes? *Journal of Substance Abuse Treatment*. 2008; 35(3): 223–231. [PubMed: 18248942]
- Masclot D, Noussair C, Tucker S, Villeval MC. Monetary and non-monetary punishment in the voluntary contributions mechanism. *American Economic Review*. 2003; 93(1):366–380.
- McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. *Annual Review of Sociology*. 2001; 27:415–444.
- Miller S, Sees C, Brown J. Key Aspects of Psychological Change in Residents of a Prison Therapeutic Community: A Focus Group Approach. *Howard Journal of Criminal Justice*. 2006; 45(2):116–128.
- Molm LD, Collett JL, Schaefer DR. Building solidarity through generalized exchange: A theory of reciprocity. *American Journal of Sociology*. 2007; 113(1):205–242.
- Nowak MA, Roch S. Upstream reciprocity and the evolution of gratitude. *Proceedings of the Royal Society B*. 2007; 274(1610):605–610. [PubMed: 17254983]

- Patenaude, A. A qualitative exploration into a prison substance abuse treatment program: "I tell them what they want to hear". In: Sims, B., editor. Substance abuse treatment with correctional clients: Practical implications for institutional and community settings. New York: Haworth Press; 2005.
- Pearce S, Pickard H. How therapeutic communities work: Specific factors related to positive outcome. *International Journal of Social Psychiatry*. 2012; 59(7):636–645. [PubMed: 22820178]
- Pedhazur, EJ., Schmelkin, LP. Measurement design, analysis: an integrated approach. Hillsdale, N.J: Lawrence Erlbaum Associates; 1991.
- Perfas, F. Deconstructing the therapeutic community: A practice guide for addiction professionals. Red Hook, NY: Hexagram Publishing; 2012.
- Phillip G. "It's just black, white or hispanic": An observational study of racializing moves in California's segregated prison reception centers. *Law & Society Review*. 2008; 42(4):735–770.
- Rand DG, Nowak MA. Human cooperation. *Trends in Cognitive Sciences*. 2013; 17(8):413–425. [PubMed: 23856025]
- Rivera MT, Soderstrom SB, Uzzi B. Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms. *Annual Review of Sociology*. 2010; 36:91–115.
- Ronel N, Elisha E, Timor U, Chen G. What do our clients say? Residents' perceptions of recovery in Retorno—A Jewish therapeutic community. *Addiction Research & Theory*. 2013; 21(4):295–305.
- Rutte C, Taborsky M. The influence of social experience on cooperative behaviour of rats (*Rattus norvegicus*): Direct vs generalised reciprocity. *Behavioral Ecology and Sociobiology*. 2008; 62(4): 499–505.
- Salter, A. Predators: Pedophiles, rapists and other sex offenders. New York: Basic Books; 2004.
- Stanca L. Measuring indirect reciprocity: Whose back do we scratch? *Journal of Economic Psychology*. 2009; 30(2):190–202.
- Stevens, A. Offender rehabilitation and therapeutic communities : enabling change the TC way. London: Routledge; 2013.
- Tiburcio NJ, Kressel D. An Enhanced Understanding of Therapeutic Communities Worldwide. *Revista Addiccion y Ciencia*. 2011; 1(4):1–18.
- Trammell R. Values, rules and keeping the peace: How men describe order and the inmate code in California prisons. *Deviant Behavior*. 2009; 30(8):746–771.
- Vandeveldel S, Broekaert E, Yates R, Kooyman M. The development of the therapeutic community in correctional establishments: A comparative retrospective account of the 'democratic' Maxwell Jones TC and the hierarchical concept-based TC in prison. *International Journal of Social Psychiatry*. 2004; 50(1):66–79. [PubMed: 15143848]
- Warren KL, Doogan N, De Leon G, Phillips GS, Moody J, Hodge A. Short-run prosocial behavior in response to receiving corrections and affirmations in three therapeutic communities. *Journal of Offender Rehabilitation*. 2013; 52(4):270–286. [PubMed: 23935258]
- Wasserman, S., Faust, K. Social network analysis: methods and applications. Cambridge; New York: Cambridge University Press; 1994.

Table 1

Mathematical formulas describing the independent variable types in the model.

Independent Variable Type Name	Formula
Reciprocity	w_{jit}
Transitivity	$w_{iht} w_{hjt}$
Indegree Activity	$w_{.it}$
Indegree Popularity	$w_{.jt}$
Outdegree Activity	$w_{i.t}$
Outdegree Popularity	$w_{j.t}$
Concentration	w_{ijt}
Isolated	$I(w_{i.t} = 0)$
Similarity on a continuous covariate	$\frac{\Delta x - x_i - x_j }{\Delta x}$
Sameness on a categorical covariate	$I(x_i = x_j)$
Covariate based activity	x_i
Covariate based popularity	x_j

w is the 3-dimensional array capturing the weight of historical events. The first subscript is the event sending actor, the second is the receiving actor, and the third is the day. A dot (.) signifies that the term is a sum across all actors on the dotted dimension.

Subscripts i and j for covariate x refer to the interaction-sending resident and receiving resident, respectively.

The I function evaluates to 1 if the contained expression is true and 0 otherwise

x represents the range of the covariate x

Table 2

Description of residents in the data, and hypotheses oriented time lagged network independent variables in the generalized linear mixed models.

	n	Mean/Proportion	sd	median	min	max
Resident Characteristics (both models)						
<i>Race -- European American</i>	1305	0.72	–	–	0	1
<i>Age</i>	1305	28.78	9.25	26	18	61
<i>Level of service inventory - Revised</i>	1305	28.14	7.18	28	5	47
<i>Length of Stay</i>	1305	140.13	36.50	149	0	184
Affirmation Model Hypothesis Variables						
<i>Res. affirmation-indegree affirmation-activity</i>	60520	1.22	0.74	1.19	0.00	4.52
<i>Res. correction-indegree affirmation-activity</i>	60520	0.73	0.63	0.70	0.00	3.66
<i>Staff affirmation-indegree affirmation-activity</i>	60520	0.23	0.44	0.00	0.00	5.88
<i>Staff correction-indegree affirmation-activity</i>	60520	0.15	0.37	0.00	0.00	5.54

Table 3

Results of the generalized linear mixed model of affirmation events.

	b	95% CI Lower Bound	95% CI Upper Bound
Unit Level Fixed Effects			
Rural unit 1	-1.949	-2.294	-1.634
Rural unit 2	-2.307	-2.681	-1.976
urban unit 1	-3.053	-3.409	-2.702
urban unit 2	-2.029	-2.344	-1.723
Focal Hypothesis Model Parameters			
affirmation-indegree affirmation-activity	0.177	0.123	0.228
correction-indegree affirmation-activity	0.045	-0.012	0.103
Staff affirmation-indegree affirmation-activity	0.035	-0.028	0.104
Staff correction-indegree affirmation-activity	-0.125	-0.197	-0.050
Ancillary Hypothesis Model Parameters			
reciprocity	0.774	0.614	0.941
transitivity	-0.008	-0.114	0.093
affirmation-indegree affirmation-popularity	0.125	0.066	0.189
affirmation-outdegree affirmation-activity	0.409	0.374	0.445
affirmation-outdegree affirmation-popularity	-0.004	-0.041	0.033
affirmation concentration	0.237	0.099	0.372
affirmation isolation	-0.626	-0.723	-0.534
cross-network reciprocity	-0.073	-0.347	0.209
correction-outdegree affirmation-popularity	0.028	-0.019	0.076
cross-network affirmation concentration	0.310	0.088	0.538
correction-indegree affirmation-popularity	-0.014	-0.075	0.044
correction-outdegree affirmation-activity	0.136	0.087	0.181
Time similarity (cohort homophily)	2.569	2.178	2.937
First month activity	-0.192	-0.274	-0.106
First month popularity	0.044	-0.026	0.114
> 4 months activity	-0.337	-0.418	-0.259
> 4 months popularity	-0.208	-0.276	-0.139
age similarity	0.331	0.149	0.520
same race	0.333	0.265	0.406
LSI-R similarity	0.003	-0.222	0.205
age activity	0.002	-0.004	0.009
age popularity	0.004	-0.001	0.008
European American activity	-0.083	-0.220	0.058
European American popularity	-0.119	-0.216	-0.028
LSI-R activity	-0.015	-0.024	-0.006
LSI-R popularity	-0.001	-0.007	0.005
Random Effect Variances			

	b	95% CI Lower Bound	95% CI Upper Bound
sender	0.782	0.726	0.840
receiver	0.385	0.349	0.427
Day	1.333	1.205	1.484
Other			
minimum-deviance decay factor δ	44313.491	44165.293	44475.186
model deviance	0.71		