# A Social Network Model of Direct versus Indirect Reciprocity in a Corrections-Based Therapeutic Community


*Nathan Doogan, Keith Warren, Danielle Hiance, Jessica Linley*

*The Ohio State University*


## Abstract


Therapeutic communities (TC) for substance abuse depend on the mutual aid among a group of residents to positively support the recovery process. An example of this aid arises from the expectation that TC residents will provide affirmations to each other for actions that are considered positive toward recovery or beneficial to the resident community. TCs are effective at treating substance abuse, however, it is unclear how the cooperative behaviors among residents are maintained. In light of a recent call for TC process studies, this paper reports on a novel agent-based computer model in which affirmations are recorded as directional arcs in a social network. Reception of affirmations by agents can be modeled based on either *direct reciprocity* (DR) or *indirect reciprocity* (IR). Respectively, an agent focuses its reciprocal actions towards those from whom it has received affirmations, or an agent focuses its reciprocal actions towards other agents that have a reputation for sending affirmations. The resulting modeled social networks are then compared to the social networks found in two Ohio TCs. This study finds that IR more closely mimics the level of reciprocity and transitivity found in the TC network than does the DR model. The findings of this exploratory study suggest the mechanisms underlying DR are unlikely to explain the interactions among TC residents. Also of interest is that social network tools may have a useful place in the day-to-day operation of TCs.


## Introduction

• Residential therapeutic communities (TC) are the most common form of substance abuse treatment in the American correctional system.

• Such TCs house as many as 150 residents at one time.

• TCs are based on mutual aid between residents, who are expected to support each other in recovering from substance abuse.

• TC researchers have gone so far as to say that the *community* of fellow substance abusers *is the method* of treatment.

• Qualitative studies have suggested that TC residents do value the help of their peers, but take several months to internalize the TC culture of mutual aid.

• At this point there has been no quantitative study of the maintenance of cooperation in a TC.

• This study aims to address the lack of research using a social network perspective of resident interactions to explore the mechanisms supporting cooperation among the residents of a TC.

## Theory

### GAME THEORETIC MODELS OF COOPERATION IN GROUPS

Direct Reciprocity: In a computer-based simulation tournament, Axelrod and Hamilton (1981) found that a tit-for-tat model of direct reciprocity was able to produce cooperation in groups if the chances of meeting another player more than once during a multi-round game is sufficiently high.

I will cooperate with peers with whom I have experienced cooperation before.

## Theory (continued)

Indirect Reciprocity: Consists of two subtly distinct phenomena: *upstream* and *downstream* indirect reciprocity.

Upstream: If someone scratches my back, I will be more likely to scratch someone's back (pay it forward).
Downstream: If you scratch someone's back, I will be more likely to scratch your back (reputation).

Martin Nowak (2005) compares *direct reciprocity* to barter. *Downstream indirect reciprocity* could be compared to behavior regarding internet shopping sites. The number of online stores is so large that a relationship between an average buyer and the store is unlikely. But I will buy from a particular site if I am sufficiently convinced by their reputation that I am not going to get taken.

### DIRECT AND INDIRECT RECIPROCITY IN THE TC

TC research suggests that *upstream indirect reciprocity* is a relevant construct in the study of TCs as affirmations are known to motivate residents to help others (DeLeon, 2000).

It is less clear what role *direct* and *indirect reciprocity* play in the cooperative behaviors of residents in a TC. Affirming good behavior could be thought of as a risky venture among this population. A rejection could be relatively costly. It would stand to reason that a resident would do well to have a prior relationship with another before affirming behaviors. This matches the model of *direct reciprocity.*

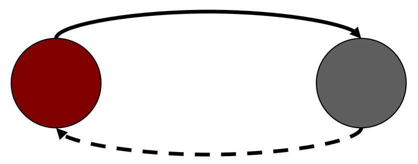
On the other hand, TC theory suggests that role models ought to be held in high regards in the community. Therefore, staff encourage and reward affirmative messages among residents. Additionally, the senders and receivers of affirmative messages are often announced at meal times in the TC. Such activity could potentially lead to a reputation for helpful behaviors. This matches the model of reputation – *indirect reciprocity.*

## Method

### AGENT BASED MODELS

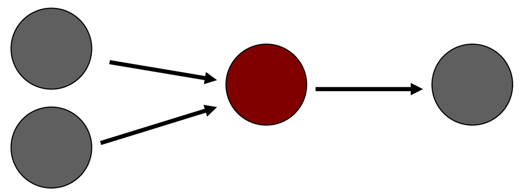
For this study, two agent-based models (ABM) are developed to simulate the sending and reception of affirmations in a TC: one based on *direct reciprocity (DR),* the other based on *downstream indirect reciprocity (IR)* (reputation). Both models utilize *upstream indirect reciprocity* to model the likelihood of an agent sending an affirmation. DR and IR are the mechanisms that determine who will receive the sent affirmation. DR bases this choice on previous interactions. IR bases this choice on the reputation of other agents for sending affirmations.

Direct Reciprocity (DR):



In the DR model, agents keep a list of other agents from whom they have received an affirmation. Agents select a receiver from this list. Each agent has one random agent placed on their list each round to get interaction started. Once the lists grow, the probability of the random agent being chosen drops.

Indirect Reciprocity (IR):



Affirmation receivers are selected probabilistically from the entire population based on the number of affirmations they have sent in the past.

Both models are run over many iterations of one of these two receiver selection processes.

## Method (continued)

### ANALYSIS

Data: Data of sent affirmations has been captured from real resident interactions in two Ohio TCs (NW and WC). Affirmations in the TCs are recorded and vetted by staff and residents for sincerity. The IDs of affirmation senders and receivers are stored in a database. These records can be represented as a social network.

The ABMs also keep track of the IDs of sending and receiving agents. This list is identical in format to the data captured from the TC and therefore can also be represented as a social network.

Model Fitting: ABM parameters are tuned to known values of network size and density of network ties. A parameter controlling the amount of effect the reception of an affirmation has on the motivation of the receiving agent to send an affirmation is the only parameter that was varied to fit the models to empirical data.

The social networks produced by the TC ABMs and by the residents of the real TC are compared on four social network measures for similarity – two measures of direct reciprocity, transitivity and triangle completion. Transitivity is a measure of network closure that can be regarded as a measure of agreement by residents about who should be affirmed. Triangle completion is another measure of network closure that implies indirectly reciprocal interactions.

## Results

Table 1 shows measurements of TC and model networks. The DR model could not be tuned to produce measurements that resemble the TC network. The IR model was able to be tuned to produce networks with a higher degree of similarity to the TC networks.

Table 1. Measurements of ABM and TC networks for comparison.

| Measure | DR<br>NW Only (settings comparable to IR) | IR<br>WCCCF / NW | Empirical Data<br>WCCCF / NW |
|---|---|---|---|
| Dyad-based Reciprocity | .609 | .380 / .308 | .388 /.299 |
| Arc-based Reciprocity | .757 | .551 / .471 | .559 /.460 |
| Transitivity | .164 | .453 / .368 | .470 / .383 |
| Triangle Completion | .060 | .216 / .161 | .213 / .155 |

Table 2. Measurements of DR ABM networks given a sweep of the affirmation reception effect size parameter. NW TC network measurements included for comparison.

| | Affirmation Reception Effect | | | | | |
|---|---|---|---|---|---|---|
| **Network Measures** | **0.50** | **0.90** | **1.10** | **1.50** | **2.00** | Empirical - NW |
| Dyad Reciprocity | 0.622 | 0.593 | 0.609 | 0.717 | 0.755 | 0.299 |
| Arc Reciprocity | 0.767 | 0.745 | 0.757 | 0.835 | 0.860 | 0.460 |
| Transitivity | 0.169 | 0.170 | 0.164 | 0.141 | 0.134 | 0.383 |
| Triangle Completion | 0.062 | 0.062 | 0.060 | 0.051 | 0.049 | 0.155 |

These tables indicate that the DR model tends too strongly towards reciprocal ties to match the TC data regardless of parameter settings. Relative to the IR model, the DR model produces far too many directly reciprocated ties and far too few transitive triads and completed triangles. The IR model, on the other hand, appears to match the TC data much more closely.

## Discussion

• The preceding analysis is exploratory

• Only two empirical networks examined

• Statistics are descriptive network stats

• Reports are of individual simulations (i.e. no probability tests). Therefore, while the IR model appears to closely match the data, it is not clear that the IR mechanism is analogous to the mechanism underlying the evolution of the real TC social network.

• The models are likely to be too simple to capture the effects of diversity among the residents in the TCs or appropriately capture changes in individual behaviors that are theorized by TC researchers.

• Important to note about model fitting

• The DR and IR models were not programmed to produce a certain number of triads or reciprocal dyads. The results of these global measurements arose from local interactions of agents that are based on the mechanisms described in the *method* section of this poster.

• Despite limitations, it appears *direct reciprocity* is not likely to be the only consideration at the individual level regarding who a resident will affirm. *Indirect reciprocity,* on the other hand, appears to be worthy of future research.

• The DR model could not reproduce the network structure observed in the TC.

• The IR model did a relatively much better job mimicking the TC network on the measured dimensions.

• This has some implications regarding the daily activities in the TC as well as for future research.

• Perhaps reputation through indirect means – such as being present during an affirmation, or by hearing an affirmation announcement at lunch-time – plays a role in resident decisions regarding to whom an affirmation should be sent. TC researchers and practitioners should be aware of this as future research unfolds.

• Future analysis of social networks in TCs might involve a series of simulations such that confidence intervals can be formed around the parameter values, thus allowing hypothesis testing to be completed.

• Alternative methods of analysis could include ERGM / P* models or longitudinal analogs such as SIENA to model the evolution of the social network over time with considerations for individual diversity among the actors in the network. These methods use well defined estimation procedures, greatly simplifying the process of developing an appropriate model to fit the data.

• Understanding what makes one TC more effective than another has important implications for the future of TC development and maintenance.

While these results can not speak directly to what makes one TC more effective than another, it certainly breaks ground for such study to begin. The aforementioned methods, given appropriate data, are capable of determining the micro-level mechanisms underlying social behavior in TCs – successful or not. Further study of the social, political, and physical components of TCs that lead to such mechanisms would make a revolution in the understanding of TC process seem to be just around the corner.