Social ties formation in an evolutional model

Rita Maestri

June 19, 2019

1 Abstract

The aim of the project is to simulate the formation of conversational groups in the context of a room full of people, and then to study the formation and evolution of the social network that emerges from the interaction within agents.

2 Introduction

It is observed that during an encounter occasion (parties, classes, etc.) people tend to form conversational groups that vary in number and composition of partecipants. But how do conversational groups form? And what kind of social network emerges from theese encounters?

Theory of conversational groups

Some of the most important variables that influence the decision of beginning a conversation with a stranger are, for example, demographic information such as color and gender. But the main factor that causes the beginning of a conversation is proxemity.[1]

In fact, there are several studies about the generation and the spatial organisation of conversational groups, which have been called *F-formation*. "An F-formation arises whenever two or more people sustain a spatial and orientational relationship in which the space between them is one to which they have equal, direct, and exclusive access."[2] "An F-formation is formed when two or more people adopt appropriate interpersonal distances and face one another, such that their individual transactional segments overlap, leading to the emergence of an overlapped space called the o-space."[3]

When the transactional segments of two or more people overlap, there is an F-formation, because those people face one another and the distance between

[1] R. E. Sykes, Initial interaction between strangers and acquaintances, 1983

[2] F. Setti, C. Russell, C. Bassetti, M. Cristani - *F-Formation Detection: Individuating Free-Standing Conversational Groups in Images*, PLoS One. 2015

[3] Kavin Preethi Narasimhan - *Computational Proxemics: Simulation-based analysis of the spatial patterns of conversational groups*, 2016

them is the social distance (which is the distance that people usually converse at).

Theory of social networks

Network studies show that real networks exhibit some recurrent characteristics.

- *Clustering coefficient.* The clustering coefficient of a network is the average fraction of pairs of neighbors of a node that are also neighbors of each other. It has been found that most large-scale real networks have a tendency toward clustering, in the sense that their clustering coefficients are much greater than the clustering coefficient of random networks $< C_R >= \frac{2<L>}{N(N-1)}$, where N is the number of nodes and $< L >$ is the mean number of links created in a random graph with a given probability of developing a link between any two nodes.[4]

- *Degree distribution.* It has been observed that, in many real networks such as social networks, the nodes degree distribution follows a power law (and is called scale-free for this reason), while for random graph it is similar to a Poisson distribution.

Hypothesis

In the simulation, all the agents are strangers to each other at the beginning. At first, they begin conversation with strangers, but after a while they are allowed to join a conversational group only if there is at least a person they already know in the group.

My hypothesis is that agents which have gained more friends at the beginning of the simulation (possibly the ones with a larger propensity to talk with strangers) are going to gain even more friends. In fact, they will be able to join more conversational groups since they know more agents. This is the "rich gets richer" idea that gives place to scale free degree distribution.

Moreover, the friendship network created during the simulation is expected to have a greater clustering coefficient than the one of a random graph with the same number of links.

In closing, an important finding in sociological reaserch[5] is that usually the conversational group size does not exceed 4 people. It is interesting to verify if this characteristic emerges from the simulation.

3 Method

In the simulation, both proxemity and demographic information is taken into account to make the agent choose whether or not to start a conversation with another agent. Moreover, the choice depends on two additional variables: propen-

[4] X. F. Wang and G. Chen, *Complex Networks: small world, scale-free and beyond*, 2003

[5] R. I. M. Dunbar - *size and structure of freely forming conversational groups*, 1995

sity to talk with strangers and the friendship link that already exists between the agents.

NetLogo software is used to perform the simulation. Three main challenges come up in the program: the choice of agents' attributes, social network creation and modelling, and the conversational group formation implementation.

3.1 Agents

All the agents have several characteristics.

- Innate attribute, which the probability of beginning a conversation depends on. They are, for example:
  - the propensity to interact with stangers;
  - the gender: it is more probable that two agents begin a conversation with one another if they have the same gender;
  - a home, wich is one of the four patches at the angles of the world: the agent has a probility of 0.1 to move towards its home when it changes direction. This way we expect to have four clusters of agent that will meet each other more often.
- Dynamic attributes that determine their status, for example:
  - a boolean attribute that indicates whether the agent is in a conversation or not;
  - the time spent in the conversation, which is a counter of the time the agent is spending in the same group of conversation.

3.2 Social Networks

Two dynamic networks are generated in the program: friendship and conversation networks.

Friendship network

Frendship network has the agents as nodes, and the link between two nodes have a weight different form zero if they have had at least one conversation together.

At the beginning of the simulation, all the weights of friendship link are initialized as zero, so the agents are all strangers to each other. The weights w increase their value as a function of the time spent in a conversation together t:

$$w = 1 - \exp \alpha t \tag{1}$$

The properties of this network at the end of the simulation are the main object of my analysis (see section 4).

Conversation network

Conversation network has the agents as nodes, and the link between two agents are active only if they are conversing.

Conversation network is useful for the visualization of the conversations during the simulation.

It is also interesting to study how many conversation networks are generated during the simulation, and how many agents were involved in the conversation.

3.3 Implementation of agent's movements and conversational group formation

The agents move randomly through the world, but there is a probability $P = 0,1$ that when they change direction of motion they start moving towards their home. We then expect that agents with the same home are going to have more occasions to converse, because they are going to be close more often.

The detailed description of the conversational group formation is shown below.

- **Encounter.** To simplify the implementation, the intersection of transactional segments is approximated in the program: two agents may begin a conversation if the first is in the cone of view of the second and viceversa (this way they face each other), where this cone has predefined lenght and angle. We have two cases.

  1. The two agents are not friends (the weight of the friendship link is zero). In this case the probability to start talking to the other agent has an offset, given by the propensity to speak to strangers and whether they have the same gender or not. The probability increases with the number of times the two agents have met before but is inversely proportional to the number of friends the agents already have (this is because if one has no friends will have a tendency to be more sociable with strangers).
  2. The two agents already know each other: the probability of beginning a conversation has an offset which is the same as the previous one (this way it will be more probable to talk with friends than with strangers), and it is proportional to the weight of the friendship link between the two agents.

- **Inclusion of other agents in the conversation.** If a new agent is facing one of the agents in a conversation, we have two cases.

  1. The agent who wants to be part of the F-formation doesn't know anyone in the F-formation. In this case, the probability of accessing the conversation is proportional to the propensity of speaking with strangers of the candidate and to the mean of the number of times the candidate has met each one of the conversing agents. It is inversely proportional to the number of friends the candidate already has.

2. The agent who wants to be part of the F-formation knows at least one agent in the F-formation: in this case the probability to join the conversation is the mean of the strenght of the relationship with the members of f-formations.

- **During the conversation.** In this phase, the weights of the friendship network are updated, as a function of the time spent in the conversation (see equation 1).
- **Leaving the F-formation.** Each agent decides to leave the conversation with a probability proportional to the residence time in the conversation.

The simulation

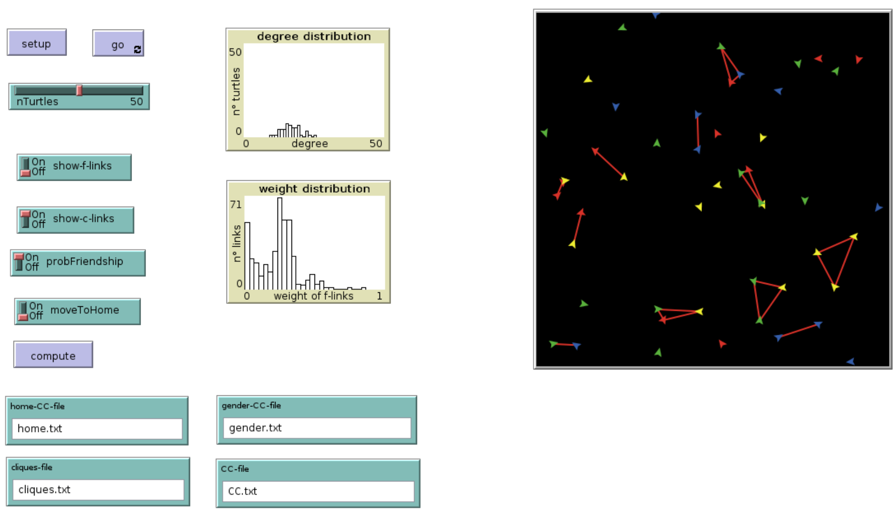

Figure 1: User interface of the simulation. In the world view, the visualisation of conversation links is active.

The simulation lasts 6000 ticks. All the weights of friendship network are initialised as zero. Until the mean degree of friendship network degree distribution equals $< D >= \frac{number\ of\ agents}{10}$, the probability to interact with other agents depends only on innate attribute (see case 1. of *encounter* and *inclusion of other agents* description, in section 3.3). When the mean degree excedes this quantity, the same probability depends on the weight of friendship link (see case 2. of *encounter* and *inclusion of other agents* description, in section 3.3). This way agents become friends with unknown ones at the beginning, but later this becomes possible only through a common friend that allows the agent to join the conversation.

As it can be seen in Figure 1, at each tick of the simulation the degree distribution and the friendship link weight distribution plot is updated.

It is possible to choose to show in the world screen the conversation links or the friendship links (see Figure 1 for conversation links, Figure 2 for friendship links).

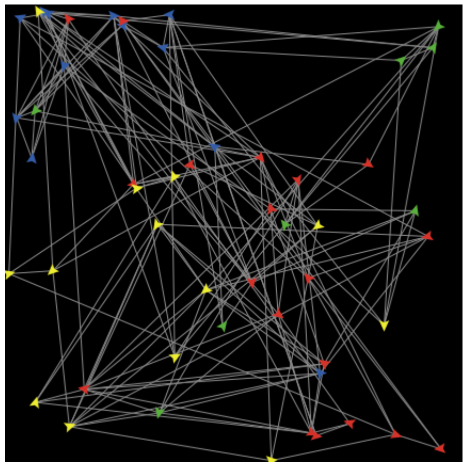

Figure 2: View of the world with friendship link visualisation active. Agents with the same color have the same home.

It is also possible to perform the simulation not involving the probability: it means that every time two agents face each other they start a conversation, not caring about the value of innate attributes or the weight of friendship link. This way it is possible to compare the results in the two cases.

4 Results

Two kinds of simulations were performed to compare the results. In the first kind (H) the agents move toward their home according to the same modality described in section 3.3. In the second kind (NH) the agents move randomly. Performing the NH simulation excluded the possibility that a higher clustering coefficient is only due to the proxemity factor. All the simulations ran for 6000 ticks. Some interesting quantities of the two networks are analised below.

- **Clustering coefficient.** The clustering coefficient of the friendship network was calculated to compare it with the one of a random network and of some subset of the network expected to have a higher CC (see Table 1). In particular, the CC of agents with the same home and with the same gender were calculated. Moreover, the evolution of clustering coefficient was recorded and compared with the one af a random network with the same number of links, in the case of NH simulations.

Table 1: Mean CC for H simulations

number of agents	friendship network CC	random network CC	CC by gender	CC by home
30	$0,83 \pm 0,01$	$0,83 \pm 0,01$	$0,87 \pm 0,03$	1
50	$0,76 \pm 0,01$	$0,75 \pm 0,01$	$0,81 \pm 0,02$	$0,95 \pm 0,02$
70	$0,72 \pm 0,01$	$0,71 \pm 0,02$	$0,76 \pm 0,03$	$0,93 \pm 0,03$

Table 2: Mean CC for NH simulations

number of agents	friendship network CC	random network CC	CC by gender	CC by home
30	$0,90 \pm 0,02$	$0,90 \pm 0,01$	$0,93 \pm 0,02$	$0,90 \pm 0,09$
50	$0,82 \pm 0,01$	$0,82 \pm 0,01$	$0,86 \pm 0,02$	$0,83 \pm 0,05$
70	$0,72 \pm 0,01$	$0,71 \pm 0,01$	$0,76 \pm 0,03$	$0,73 \pm 0,03$

Mean CC over 20 runs for different numbers of agents involved in the H (table 1) and NH (table 2) simulations. The error is the sandard deviation. The friendship network CC is the mean clustering coefficient for the entire network. The random network CC is calculated over a random network with the same number of links as the friendship one. The CC by gender is calculated over a subset of agents with the same gender. The CC by home is calculated over a subset of agents with the same home.

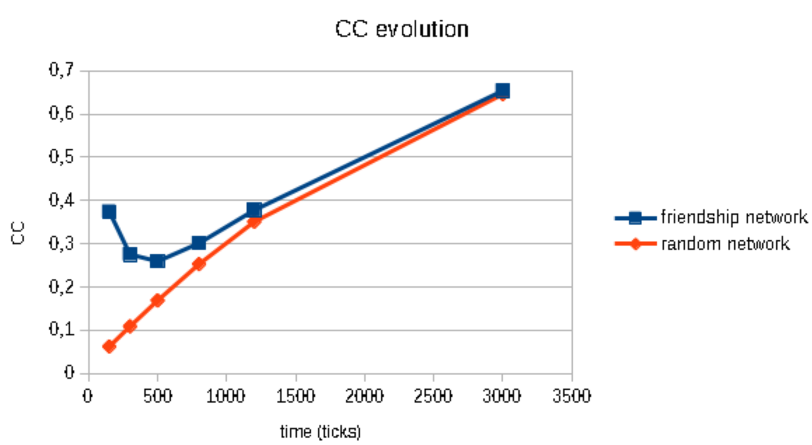

Figure 3: Mean of clustering coefficients (calculated over 20 simulations with 50 agents) of the friendship network and of the correspondent random network for different times.

It is shown in Figure 3 that the CC of the network evolved during the simulation is higher than the one of a random network only when the simulation is at the beginning, i.e. when the density of the network is lower. In fact after some time, the clustering coefficient of the simulation network asimptotically tends towards the one of a random network. This result suggests that meeting friends of friends generates a highly clusterized friendship network as long as the the links are just a few, later the links are more uniformely distributed.

It is also observed that, in H simulations, the CC of agents with the same home is higher than the one of the entire network. This means that agents with the same home tend to create more links because they meet more often. Therefore, the model is well approximating the real behavior of conversational groups, which tends to form mainly because of the proxemity within the members.

The data also suggest that including in the model a higher probability of speaking with agents with the same gender produces an effect on the network: in fact, the CC for this kind of agents is slightly higher than the one of the whole network. This result well represents the homophily observed in people interaction.

- **Degree Distribution.** The degree distribution has been observed throughout the whole duration of simulations through a dynamic graph of netlogo (see figure 1).

  It is found that the degree distribution is similar to a Poisson distribution whose mean shifts towards right as the time passes. Therefore, this model does not explain the scale free degree distribution observed in real social networks. In fact it is actually expected that, if each agent meets at least one other agent in the first part of the simulation, as time passes the probability that he doesn't meet that agent again (and, consequently, that his friend doesn't introduce him to new agents), goes to zero. This implies that every agent is going to have a growing numbr of friends, and there is no reason that

- **Number of partecipants to conversation**. The conversational group size distribution that emerges from the simulation is similar to the one found by Dunbar in his reaserch (5). In fact, both exhibit an exponential decay ($R^2 = 0,98$ for H simulations with 70 agents, $R^2 = 0,97$ for Dunbar's dataset). In fact, the model correctly predicts the observed low probability of having conversational groups which exceed the 4 people.

Table 3: Distribution of conversational groups size

Group size	Dunbar's results	simulation 30 agents	simulation 50 agents	simulation 70 agents
2	$53,9\%$	$(69 \pm 2)\%$	$(67 \pm 1)\%$	$(65 \pm 1)\%$
3	$26,9\%$	$(23,3 \pm 0,9)\%$	$(24,7 \pm 0,6)\%$	$(25,8 \pm 0,3)\%$
4	$13,5\%$	$(5,7 \pm 0,9)\%$	$(6,1 \pm 0,4)\%$	$(7,1 \pm 0,4)\%$
5	$4,7\%$	$(1,1 \pm 0,8)\%$	$(1,3 \pm 0,2)\%$	$(1,6 \pm 0,2)\%$
6	$0,7\%$	$(0,2 \pm 0,01)\%$	$(0,2 \pm 0,1)\%$	$(0,3 \pm 0,1)\%$

Distribution of conversational groups size. In the last three columns there are the mean of the percentage of groups of a specific sized over the totality of groups formed, calculated over 20 runs. The error is the standard deviation. In the second column the results found by Dunbar were reported (5).

The group size in the model depends on the agents' density, as it can be seen from the data in Table 3, but most of all it depends on the time spent in the conversation. In fact, I chose the conversation abandonment time distribution observing which distribution would give the most realistic result: for high abandonment times the conversational groups would have been too big, for short times the opposite would have happened.