Collaboration Analysis in Recommender Systems using Social Networks

Jordi Palau, Miquel Montaner, Beatriz López, and Josep Lluís de la Rosa

Institut d'Informàtica i Aplicacions
Agents Research Laboratory
Universitat de Girona
17071 Girona, Spain
{jpalaur, mmontane, blopez, peplluis}@eia.udg.es

Abstract. Many researchers have focused their efforts on developing collaborative recommender systems. It has been proved that the use of collaboration in such systems improves performance, but what is not known is how this collaboration is done and what is more important, how it has to be done in order to optimise the information exchange. The collaborative relationships in recommender systems can be represented as a social network. In this paper we propose several measures to analyse collaboration based on social network analysis. Once these measures are explained, we use them to evaluate a concrete example of collaboration in a real recommender system.

Keywords

Recommender Systems, Collaboration Analysis, Electronic Communities, Social Networks, Trust

1 Introduction

In the real world, not only society in general but in particular our friends, help us to discover new things which they think we would like. Our friends advise us about an interesting product, a movie, a book or a restaurant, collaborating with us in the selection process. Being aware of this collaboration in the real world, researchers have focused on the development of recommender systems [14] which can recommend items to a user based on information from other users.

Particularly, the collaborative filtering method has proved to be a useful method to take advantage of the collaborative world especially when combined with other technologies in a hybrid approach [2, 6]. Thus, the collaboration among users increases the performance of recommender systems. However, we do not know many things about how this collaboration is done. This is a first step towards the design of new methods and techniques that will contribute to optimise collaboration with a given purpose (goal).

Recently, collaboration has been modeled as a network of users exchanging information, that is, a social network. Users are represented as actors (nodes) and collaborative relationships as directed ties. In this paper we use this representation to propose several measures based on social network analysis in order to understand how users collaborate.

To illustrate the use of the measures, we perform the evaluation of a real collaboration framework implemented in our group.

Thus, this work is a first step to achieving a further goal. The long-term aim of our work is to find out how can we tune the different parameters of our recommender system in order to have a social network featured in an optimal way according to certain criteria.

This paper is structured as follows. Section 2 introduces social networks and why they are used in our work. Our proposal of measures to analyse collaborative recommender systems are presented in Section 3. Section 4 introduces the collaborative recommender system implemented in our group and used as a basis for our experimentation and Section 5 shows how the proposed measures are used to analyse our real example. Section 6 presents related work and finally, some conclusions and further work are provided in Section 7.

2 Social Networks

A social network [3] is a representation of the relationships existing within a community. Social networks provide us with a tool to study collaboration, in particular through theory developed in social network analysis [18, 16, 7].

Even within the same community several types of social networks can be built depending on the social relationship taken into account: friendship, mutual support, cooperation and similarity are typical criteria used in establishing the social relationship components of a community. Actors in this social network can be individuals, groups of people, objects or events as far as certain relationships hold them together. The strength of a tie may range from weak to strong depending on the quantity, quality and frequency of the exchanges between actors [12].

In this way, social networks are able to represent societies and relationships among individuals from these societies by means of a graph. In collaborative recommender systems, each system user is represented by an actor in the graph, and relationships among these users are represented through directed ties. If user A develops a relationship with user B, there should be a directed tie from A to B.

Since relations among users change over time, it is important to take into account that social networks are dynamic. So, a social network represents relationships among users at a certain moment in time.

It is also important to know which locality the analysed system has. A system with locality 1 is the one where only the immediate ties a user has are taken into account. In systems where this locality is higher than 1, immediate and also indirect ties are considered. For example, if there is a user A connected to another user B, and B is connected to a third user C, A can reach C through B. However, in systems with locality 1, A cannot reach C unless there is a tie between them.

3 Social Network Measures

This section presents the measures we propose in order to analyse the collaboration among the users/agents of a recommender system. In general, the social network resulting from a collaborative recommender system has locality 1. Therefore, the measures proposed in this paper only take into account the immediate ties among the actors. All these measures are based on social network analysis; namely *size, density, degree centrality, network centrality, clique membership* and *factions*.

3.1 Size

Size is the number of actors present in the network, and is useful in order to calculate other measures.

This parameter can give us a general idea of how the network is. Say we have a small firm with only 10 workers. It would be easy for each worker to know the others and build up relationships. Now imagine we have a firm with 1000 workers. It would be extremely difficult for any worker to know all of the others. As a group gets bigger (and *size* increases) the proportion of ties that could be present decreases, and usually partitioned groups emerge.

3.2 Density

Fully saturated networks (i.e. one where all logically possible ties are actually present) are rare, especially in social networks with a considerable number of actors. In a network whose *size* is K, the number of possible different directed ties is (K * (K-1)).

Density is the proportion of all ties that could be present that actually do in fact exist. A low *density* tells us the system analysed is restrictive when actors have to establish relationships with other actors. In the other hand, in a high *density* system, relationships amongst actors can easily be made.

3.3 Degree Centrality

Degree is a measure that counts the number of ties an actor has. In the case where we are dealing with a network where direction of ties is important, we can distinguish between *in-degree* and *out-degree*.

On one side, *in-degree* is the number of ties an actor receives. According to social network theory, if an actor receives many ties, it is often said he has high prestige because many other actors seek to direct him ties. This approach can be applied in our study because if an actor receives many ties, it means other actors trust him, so he has more prestige. He also has more power, because he can influence other actors as far as his opinions are taken into account.

On the other hand, *out-degree* is the number of ties which begin with the actor himself. Actors with a high *out-degree* are able to make many others aware of their views. If an actor has a high *out-degree* it means he trusts a high number of other agents, and so he has more chances to ask for opinion/advice. We can say actors with high *out-degree* may be in advantaged positions because they have more alternative ways to satisfy needs and they are less dependent on other actors or they may have access to more resources.

3.4 Network Centrality

This measure is similar to the previous one, but here the whole network is analysed instead of each actor. There are several ways to calculate *centrality* in the network, and each of them uses a different source that generates different rates. For example, it can be calculated by using *degree centrality*, *closeness centrality*, *betweenness centrality* and *flow centrality*. In our study we use the *degree centrality* calculated for each actor to calculate a value for the whole network. This decision has been taken because *degree centrality* is the only one which takes into account the immediate ties an actor has. The other ways to calculate *centrality* consider the indirect ties an actor has (i.e. actor 1 can reach actor 3 if there is an actor 2 connected to both of them). As we do not want to use this approach, we only use *degree centrality*.

Using *in* and *out-degrees*, an index of *network centrality* can be calculated. First of all we need to define the star network. A star network is a network where there is one actor A connected to all the other actors in the network. The others have only one tie (a connection to A). The star network is the most centralised network for any number of actors. We can express the degree of variability in the degrees of actors in our analysed network as a percentage of that in a star network of same size. A different value for *in-degree* and *out-degree* is calculated.

Another way to calculate *centrality* is by looking at the variation between the mean and the standard deviation for *in* and *out-degrees*. In a centralised network there is a high variation because there are huge differences within actors *in* and *out-degrees*, while in a network which is not centralised, variation tends to be lower.

The *network centralisation* parameter gives us an idea of the amount of concentration or centralisation in the whole network. A high value means that the network is centralised, that is, there are several actors who have a high *degree* and several other actors who have a low *degree* (in or out). A low value means the network is not centralised, so the actors have a similar degree value.

3.5 Clique Membership

The next two measures are related to the substructures which may be present in the network. Divisions of actors into subgroups can be an important aspect of social structure, and it can be important in understanding how a network as a whole is likely to behave.

The first structure we evaluate are *cliques*. A *clique* is a sub-set of a network in which actors are more closely tied to one another than to other members of the network. In real life people also tend to form *cliques* on the basis of age, gender, race and other criteria.

The *clique* definition is very strong as a *clique* is a number of actors who have all possible ties present among themselves (i.e. in terms of graphs, a maximal complete subgraph). Three actors can easily form a *clique*, so we do not consider these kinds of structures in this measure.

We calculate *clique membership* for each actor. First, all the *cliques* which are present in the network have to be found by considering only the ones with four or more actors. Then we get a *clique membership* which is the number of *cliques* on which an actor is a member.

Clique membership gives us an idea about the tendency each actor has to form substructures in the graph. The fact that several actors have a high *clique membership* indicates that probably there are communities within the social network because these actors are highly related among themselves. If actors have a low *clique membership* it will be extremely difficult to find communities in the network.

3.6 Factions

We have seen that cliques are very restrictive, because there must be all the possible ties present to form a clique. A less strict division would allow some ties between groups and also less than full density within them. So, the last measure we propose is *factions*. In network terms, it is possible to define partitions of the network by grouping together actors on the basis of the similarities in which they are tied.

Using the power of computers it is possible to search for partitions of a network into groups that maximise the similarity of the patterns of connections of actors within each group.

This method divides our network into the number of *factions* we want. The output is a set of different groups where actors are more likely to be tied to each other than with actors from other groups. This helps us to identify communities within our network.

4 Running Example

In an attempt to study the collaboration among users, our research group implemented GenialChef[1], a restaurant recommender system developed within the IRES Project[2]. GenialChef is the basis of our experimentation. The details of this implementation are extensively explained in [9]. As a summary, GenialChef is a multi-agent system that recommends interesting restaurants to its users. The agents making up this system can be grouped into service agents and personal agents (PA)(the system architecture is shown in Figure 1). The service agents provide objective information to the PAs: the restaurant server agents (RSA) provide information about restaurants and the personal agent facilitator (PAFA) acts as a broker agent and is in charge of assisting the PAs in finding other PAs. PAs provide personalised information to their users. Every user has a PA in the system, which encapsulates his/her user profile and is in charge of recommending to him/her interesting restaurants.

In order to take advantage of the collaborative world, PAs exchange information by means of two new information filtering methods: the opinion-based filtering method and the collaborative filtering method through trust. Their main idea is to consider other

[1] GenialChef was awarded the prize for the best university project at the E-TECH 2003.

[2] The IRES Project was awarded the special prize at the AgentCities Agent Technology Competition.

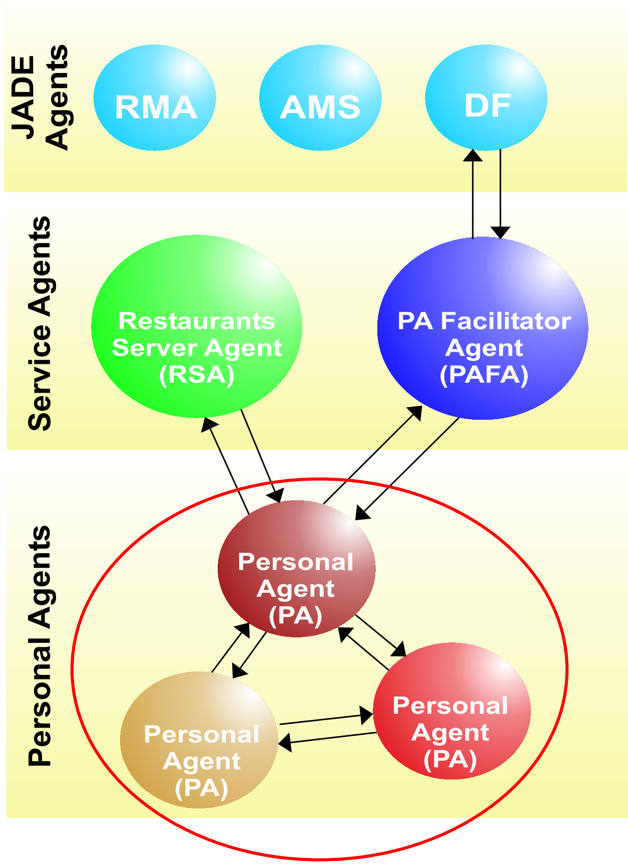

Fig. 1. System Architecture

agents as personal entities on which you can either rely or not. Thus, PAs only collaborate with reliable agents. Reliability is expressed through a trust value with which each agent labels its neighbours [10]. Once the agent has a set of reliable agents, it can use them to filter information. When an agent is not sure about a recommendation or discovers a new item, it asks the reliable agents for their opinion and uses their trust values to decide whether the item is interesting for the user or not. Moreover, PAs can ask to their friends for advice, that is, ask about new items that could be of interest to the other user. We suppose that similar agents provide pertinent opinions, but they may also give inadequate ones. Therefore, trust should be modified depending on the results of the recommendations in order to improve acquaintances.

Thus, after a period of time, each agent has a list of reliable agents with whom to collaborate in case of need *(contact list)*. Clearly, this output can be viewed as a graph representing a *social network*.

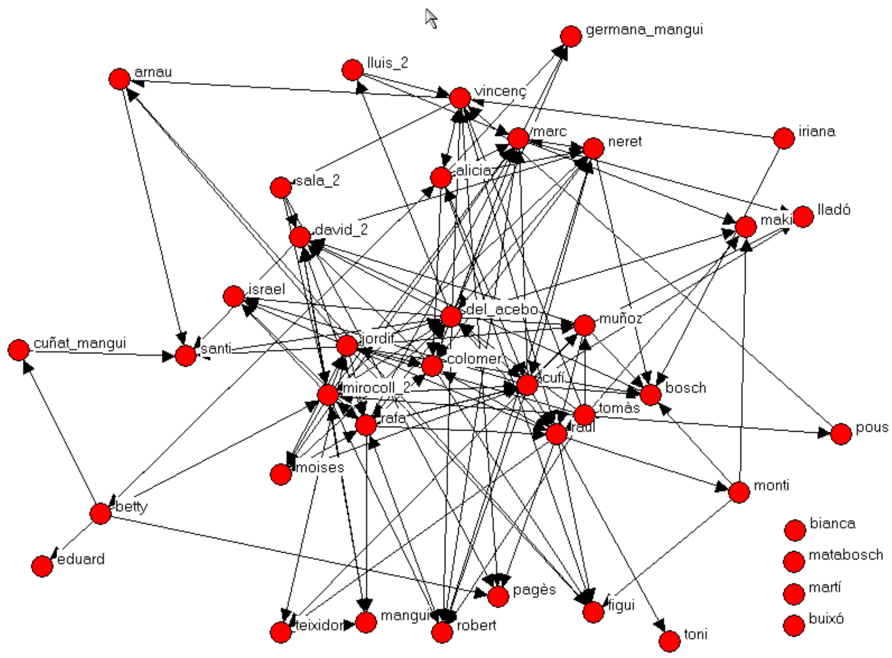

Fig. 2. Network 1: Graph layout of the relations created among PAs at the beginning of the execution, presented as a social network using Netdraw visualization program.

Therefore, each PA is represented by an actor of the graph, and the trust relationships each PA develops are directed ties among them. If a PA A trusts in another PA B, there would be a directed tie from A to B. One can find these relationships through the contact list of reliable agents that each PA has. In order to analyse the evolution of the system, two stages are considered. The first one is taken just when the system begins to execute, and the last one is taken at the end of the execution. For instance, Figure 2 and Figure 3 show the layout of the social networks obtained in our experiments at the beginning and at the end of the execution respectively. Following the example, looking at Figure 3, *sala_2* trusts *pages*, *llado* and *mirocoll* opinions/advice and collaborates in *cunat_mangui* recommendations.

It is also important to keep in mind that our social networks have locality 1, which means that only the direct ties an actor has are taken into account. For example, having three PAs A, B and C where A trusts in B and B trusts in C, does not necessarily mean that A trusts C. Therefore, some of standard techniques usually used in social network analysis can not be applied to our problem.

5 Experimental Results

We have used the measures proposed in Section 3 to analyse how the collaboration is done in a concrete framework. In particular, the collaboration among PAs from the

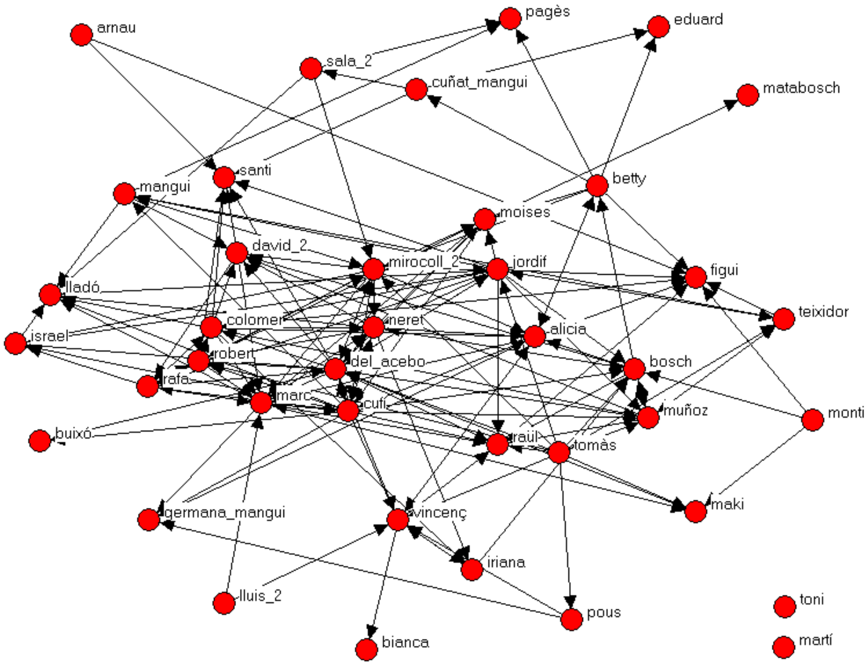

Fig. 3. Network 2: Graph layout of the relationships created among PAs at the end of the execution.

recommender system explained in Section 4 has been evaluated. In order to do that, a simulator based on the ”profile discovering procedure” has been developed [11], which allows us to perform thousands of repeatable and perfectly controlled experiments.

In this section we analyse the simulation results of a 60 day long trial with 40 real user profiles extracted from our university staff. In particular we analyse the social networks evolution, taking one picture of the social network at the beginning of the simulation (just after the startup) and another one at the end of the simulation in order to analyse the evolution of the network during the experiment. It has been demonstrated that the use of collaboration increases the performance of the system in [9]. Moreover, the performance of the system has been studied using different parameters, so we know which parameters make the system perform better. Therefore, we can extract some conclusions about what the measures obtained in the analysis of the social network should be in order to get a system with a higher performance. The simulation was performed with the optimal parameters studied in [9].

In order to evaluate the collaboration performed in this simulation, we have used UCINet [4], which is a software designed to represent and analyse social networks. In particular, we examined the different measures explained in Section 3: *size, density, degree centrality, network centrality, clique membership* and *factions*.

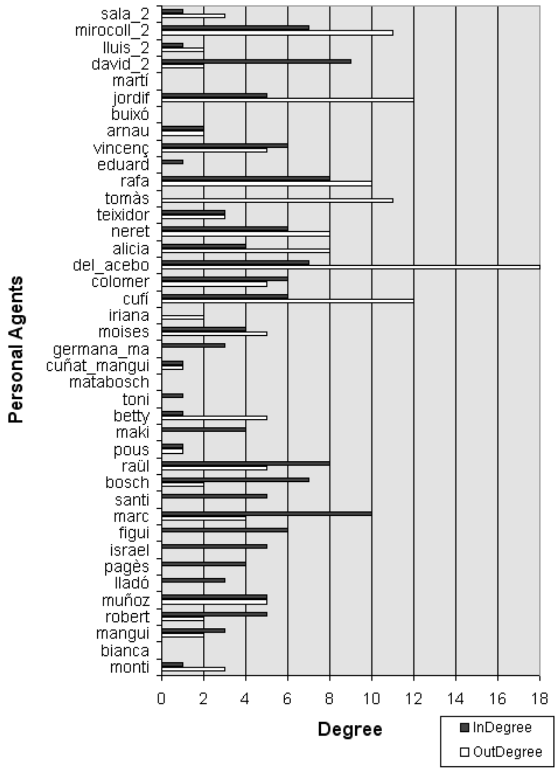

Fig. 4. Freeman's Degree at the beginning of the execution.

The first two measures to analyse are *size* and *density*. As we used 40 user profiles to run the simulation, the *size* of both resulting networks is 40. The maximum number of ties we could have in a fully saturated network of *size* 40 is 1560. In Network 1 we only have 150 real ties while in Network 2 we have 170 real ties. Both values are low and so the resulting *density* is 9.6% for Network 1 and 10.9% for Network 2. The lowness of the measures points out that, in general, PAs do not have many friends. With regard to the evolution, the *density* has increased during the experiment (there are 20 ties more at the end of the execution), which means that collaboration has allowed the PAs of the system to make new and trusted friends during the execution.

With regard to *out* and *in-degrees*, Figure 4 shows the degrees for each PA at the beginning of the execution, and Figure 5 shows the degrees after the execution. Looking at both figures, the highest *in-degree* is for `marc`, who appears on 12 other PA's contact lists. Therefore, he is the most prestigious one because there are more PAs who trust him. The highest *out-degrees* are those of `cufi`, `del_acebo` and `jordif`. They have the highest number of PA's on their contact lists to ask for opinions/advice. If we take a look at Figures 2 and 3, we see that all these PAs are drawn in a position which is quite central (in the centre of the layout). Another thing to pay attention to is the fact that there

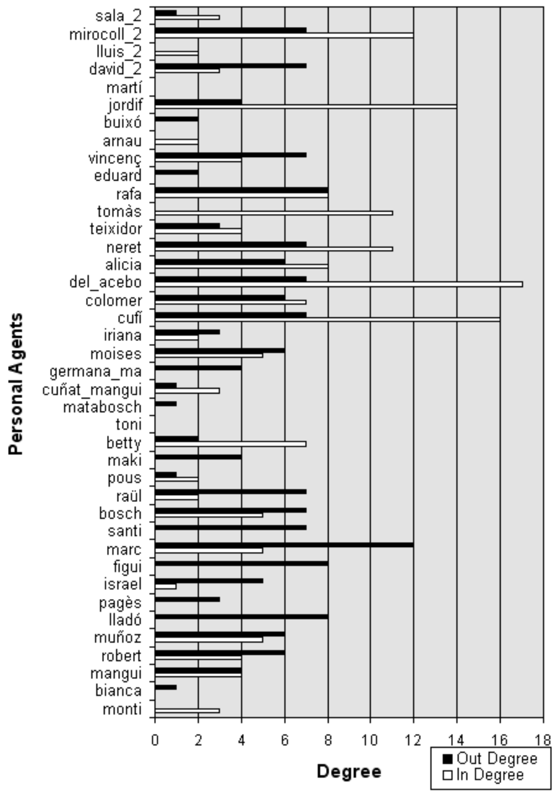

Fig. 5. Freeman's Degree at the end of the execution.

is a large number (13 at the beginning and 12 at the end) of PAs with an *out-degree* of 0. This means they do not have any PA on their contact lists and they cannot ask for advice in the case of needing it. There are also 6 PAs at the beginning and 7 at the end with an *in-degree* of 0. This means they do not have any PAs who trust them. This is due to the fact that the profiles associated with the concrete PA are different from all the others. In the real world, there are people who like restaurants that we would never go to. If we asked for advice from people with different tastes that differ from ours and we got a bad recommendation, we probably would not trust them anymore.

By analysing the evolution of these *degrees* it is possible to have an idea of the effects of collaboration in the system, and which agents are the most *collaborative*. For example, `cufi` trusted 12 other PAs at the beginning and, by the end, this number had increased to 16, so through collaboration a *friendship* had been developed with 4 new PAs. On the other hand, `llado` had 3 other PAs who trusted him at the beginning and 8 at the end, so this demonstrates that `llado` is likely to develop new ties with other PAs who start trusting him during the execution of the system. So what the analysis of *degrees* shows is which are the most and the least collaborative agents. It has been demonstrated that a system where the agents have too many friends does not perform better, and neither does a system with a low number of friends.

Having analysed *degrees* for individual PAs, we can now analyse the *degrees* in the whole network. *Network centrality* for *out-degree* is 37.54% at the beginning and 33.53% at the end and for *in-degree* is 16.5% at the beginning and 20.38% at the end. Therefore, these values are quite low, and this result would indicate that the network is not very centralised. However we have to be aware of the fact that network density is very low and there are very few ties in the network. As a consequence this result may be altered. Table 1 shows the values of *means, standard deviations* and the *coefficients of variation* for each situation. Clearly the population is more homogeneous with regard to *in-degree*, but the fact is that both values are high, so it can be concluded that structural positions are heterogeneous and that *network centralisation* is high in both networks. By comparing Network 1's results with Network 2's ones, there is a reduction of the *coefficients of variation* in the system after the execution. As a consequence, a homogeneous system should perform better than a heterogeneous one.

	Network 1		Network 2	
	In-Degree	Out-Degree	In-Degree	Out-Degree
Mean	3.72	3.72	4.25	4.25
Standard Deviation	2.81	4.33	3.03	4.61
Coefficient of Variation	75.68%	116.24%	71.32%	108.56%

Table 1. Comparison of degree means, standard deviations and coefficients of variation for both networks

Now we analyse possible network substructures. Figure 6 shows the number of *cliques* of which each actor is a member. As there are several actors who have a high *clique membership* we can say, with a reasonable certainty, that there exists at least one community in our network. There are 19 actors in Network 1 and 18 in Network 2 who have a 0 *clique membership*. This indicates that there are some actors who clearly do not belong to any community. The evolution shows that *clique memberships* at the beginning are similar to the ones at the end. The most changeable PAs regarding *clique membership* are the ones mentioned in the *in* and *out-degree* analysis, because the ones with the higher *degree* changes are the ones with the higher *clique-membership* changes.

In order to corroborate the hypothesis we came up with when analysing *clique membership*, network substructures are examined by means of *factions*. This analysis has only been made in Network 2 because the most relevant one is the resulting network at the end of the execution. UCINet's output shows that the best partition dividing the PAs in two *factions* is:

Faction 1 *moises, bosch, mangui, robert, munoz, lladó, neret, israel, tomàs, marc, santi, david_2, raül, colomer, maki, teixidor, mirocoll_2, rafa, cufí, jordif, del_acebo*
Faction 2 *betty, bianca, iriana, monti, pous, pagès, alicia, figui, matabosch, eduard, vicenç, arnau, buixó, germana_mangui, martí, toni, lluis_2, cunat_mangui, sala_2*

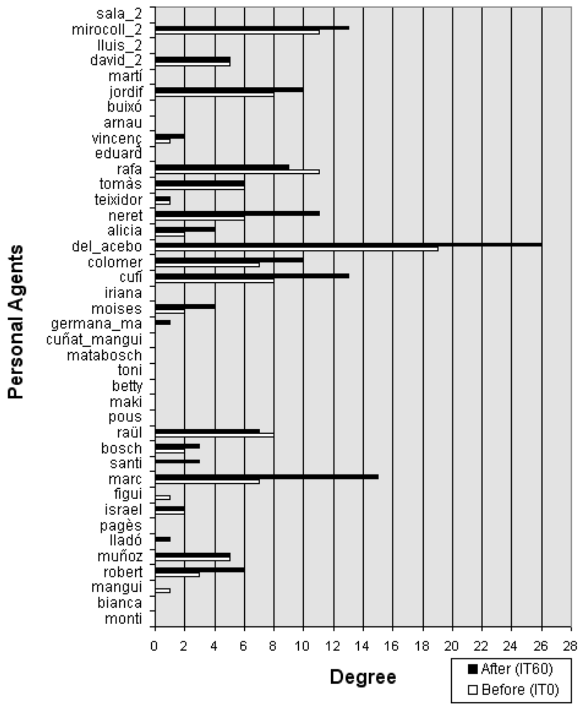

Fig. 6. Personal agents membership to cliques of size 4 or more.

We can see the distribution of ties inside and outside of the *factions* created in Figure 5. We can calculate the density of ties in the four different areas of Figure 7 in order to evaluate them. In region 1-1 the density is 0.28, in region 2-2 it is 0.06, and in regions 1-2 and 2-1 the densities are 0.05 and 0.04 respectively. On one hand there is region 1-1 (*faction 1*) with a huge concentration of ties compared to the others. This *faction* creates a community, because its PAs have developed a large number of relationships among themselves and as a consequence, they are the ones who collaborate the most. On the other hand, *faction 2* has a low density. In fact, regions 1-2 and 2-1 have almost the same density as region 2-2 (*faction 2*). Therefore, *faction 2* cannot be considered a community. This is the consequence of not having enough PAs in the network to form other communities.

6 Related Work

Social network analysis has been largely applied to other domains with different purposes. There has been a great deal of work on studying the relationships among Internet users [19]. The main objective of these studies is to find similar users on the Internet that could give useful information to others. For example, some papers use social networks in order to find communities of similar users from the Web [5, 1] or e-mail [8].

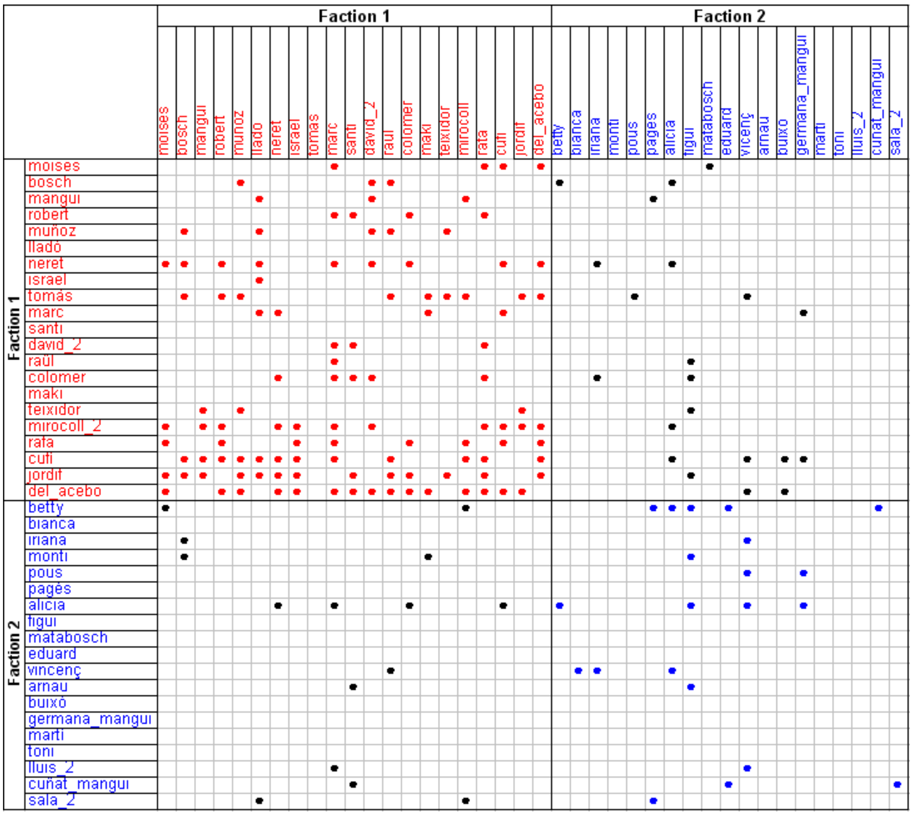

Fig. 7. Factions tie representation

Several research groups have used social networks to study trust and reputation mechanisms in multi-agent systems where agents act as assistants for the members of an electronic community. For example, some papers address the problem of calculating the degree of agent reputation needed in order to collaborate in a multi-agent system [13, 15]. Others use social networks in e-commerce to support reputations for both expertise (providing good service) and helpfulness (providing good referrals) [17].

No similar work on the utilisation of social network analysis in order to evaluate how collaboration is done has been found.

7 Conclusions and Further Work

This paper is a first attempt to analyse how users/agents collaborate in a collaborative recommender system. Up to now, efforts in research have been directed towards developing recommender systems with collaboration and demonstrating that their performance is better than the ones who do not use collaboration. The main objectives of these collaborating systems have always been focused on finding just who the best candidates to collaborate with are. However, we do not know much about how the ideal collaboration model should be in order to optimise the performance of these systems.

Our proposal in order to evaluate how collaboration is done, is by the use of social networks as a tool to represent and analyse collaboration in recommender systems. In particular we propose some measures based on social network analysis that help us to understand general aspects about the composition of the collaborative network. Thus, measures such as *size*, *density*, *degree centrality*, *network centrality*, *clique membership* and *factions* help us to achieve our objective.

Having made these measures, we used them to show how an analysis of a real system should be. In the analysed example, we were able to observe that at the end of the execution, the level of collaboration was quite low (although it was higher than at the beginning), that there was a homogeneous group of users who formed a community, and that there were several other users who were very isolated from the rest.

The results presented in this paper have been obtained through an execution of our simulator using the parameters which have been demonstrated to be optimal in [9].

Thus, the next step in our work is to perform large-scale experiments with different parameters analysing in each of them the proposed measures. In doing so, we will be able to find out the impact of the parameters of our recommender system in the social network. In particular, we want to study how the best collaboration is obtained, for example, in a centralised/decentralised network, in a dense/non-dense network or in a network with more/less communities.

Moreover, we also want to analyse how social networks evolve over time. In this paper we have only considered two snapshots of the network: one at the beginning and the other one at the end of the execution. In the future we want to study the whole evolution, and so how relationships among users are generated and how they are dropped, the social evolution of a certain user inside the social network and the progressive creation of communities and their evolution.

Acknowledgements

This work was supported by the Spanish MCYT Project DPI2001-2094-C03-01 and the Agentcities.NET Project ACNET.02.50.

References

1. L. A. Adamic and E. Adar. Friends and neighbors on the web. *Social Networks*, pages 25(3):211–230, July 2003.
2. M. Balabanovic and Y. Shoham. Combining content-based and collaborative recommendation. In *Communications of the ACM*, 1997.
3. J. Barnes. *Social Networks.* Reading, MA: Addison-Wesley, 1972.
4. F. L. Borgatti S.P., Everett M.G. Ucinet for windows: Software for social network analysis. harvard: Analytic technologies, 2002.
5. D. Gibson, J. M. Kleinberg, and P. Raghavan. Inferring web communities from link topology. In *UK Conference on Hypertext*, pages 225–234, 1998.

6. N. Good, J. Schafer, J. Konstan, A. Borchers, B. Sarwar, J. Herlocker, and J. Riedl. Combining collaborative filtering with personal agents for better recommendations. In *Proceedings of AAAI*, volume 35, pages 439–446. AAAI Press, 1999.
7. R. A. Hanneman. *Introduction to Social Network Methods*. http://faculty.ucr.edu/∼hanneman/SOC157/NETTEXT.PDF, 2001.
8. V. Krebs. Data mining email to discover social networks and communities of practice. http://www.orgnet.com/email.html, 2003.
9. M. Montaner. *Collaborative Recommender Agents Based on Case-Based Reasoning and Trust*. PhD thesis, Universitat de Girona, September 2003.
10. M. Montaner, B. López, and J. de la Rosa. Developing trust in recommender agents. In *Proceedings of the AAMAS'02*, pages 304–305, Bologna (Italy), 2002. Cristiano Castelfranchi and W. Lewis Johnson (Eds.). ACM Press.
11. M. Montaner, B. López, and J. de la Rosa. Evaluation of recommender systems through simulated users. *To be published in the 6th International Conference on Enterprise Information Systems (ICEIS'04). Porto (Portugal)*, April 2004.
12. N. Papakyriazis and M. Boudourides. Electronic weak ties in network organizations. 2001.
13. J. Pujol, R. Sanguesa, and J. Delgado. *A Ranking Algorithm Based on Graph Topology to Generate Reputation or Relevance*, volume ISBN: 3-540-44384-3, chapter 18, pages 382–395. Springer Verlag (2003), 2003.
14. P. Resnick and H. Varian. Recommender systems. In *Communications of the ACM*, pages 56–58, 1997.
15. J. Sabater and C. Sierra. Reputation and social network analysis in multi-agent systems. In *In Proceedings of the AAMAS'02*, pages 475–482, Bologna, Italy, July 15-19 2002.
16. J. Scott. *Social network analysis*. London: Sage, 1991.
17. M. Venkatraman, B. Yu, and M. Singh. Trust and reputation management in a small-world network. In *Proceedings of Fourth International Conference on MultiAgent Systems*, pages 449–450, 2000.
18. S. Wasserman and J. Galaskiewicz. *Advances in Social Network Analysis*. Sage Publications, October 1994.
19. B. Wellman, J. Salaff, D. Dimitrova, L. Garton, M. Gulia, and C. Haythornthwaite. Computer networks as social networks: Collaborative work, telework, and virtual community. *Annual Review of Sociology*, 22:213–238, 1996.