

Collaboration Analysis in Recommender Systems using Social Networks

Jordi Palau, Miquel Montaner and Beatriz López
Institut d'Informàtica i Aplicacions
Agents Research Laboratory
Universitat de Girona
17071 Girona, Spain
{jpalaur, mmontane, blopez}@eia.udg.es

Abstract

Many researchers have focussed their efforts in developing collaborative recommender systems. It has been proved that the use of collaboration in such systems improves its 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 networks analysis. Once this 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, society in general, but our friends specially, help us to discover new things which they think would like us. Our friends advise us about an interesting product, movie, book or restaurant, collaborating with us in a selection process. Being aware of this collaboration in the real world, researchers have focussed 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 an useful method to take advantage of collaborative world especially when combined with other technologies in an 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 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 are represented with 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 achieve a further goal. The long-term aim of our work is to find out which is the optimal model of social network that optimises the collaboration system.

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 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 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 network 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 individual people, groups of people, objects or events as far as certain relations 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 represent societies and the relations 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 in a certain moment of time.

It is also important to know which locality the system analysed has. A system with locality 1 is the one where only immediate ties a user has are taken into account. In system 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 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 a 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* grows up) the proportion of ties that could be present falls, and usually partitioned groups emerge.

3.2. Density

As we said, fully saturated networks (i.e. one where all logically possible ties are actually present) are rare, specially in social networks with a considerable number of actors. In a network which *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 are. A low *density* tells us the system analysed is restrictive when actors have to establish relations with other actors. On the other size, in a high *density* system relations among actors can be easily made.

3.3. Degree Centrality

Degree is a measure which counts the number of ties an actor has. In case we are dealing with a network where direction of ties is important, we can distinguish among *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 by them.

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

3.4. Network Centrality

There are several ways to calculate *centrality* in a network, and each of them uses a different method that generates a different rate. For example, there is *degree centrality*, *closeness centrality*, *betweenness centrality* and *flow centrality*. In our study we use the term *network centrality* applied to *degree centrality* regarding 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 indirect ties an actor has to others (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 each other actor in the network. All 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 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-degree*, while in a network which is not centralised, variation tends to be lower.

Network centralisation parameter gives us an idea about 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 which have a high *degree* and several other actors which have a low *degree* (in or out). A low value means the network is not centralised, so actors have a similar degree value.

3.5. Clique Membership

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

The first structure we evaluate is *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 things.

The *clique* definition is really strong, a *clique* is some 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 this kind 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, considering only the ones with four or more actors. Then we get *clique membership* as the number of *cliques* an actor is member.

Clique membership gives us an idea about the tendency each actor has to form substructures in the graph. The fact we have several actors with a high *clique membership* indicates that probably there are communities in the social network because these actors are highly related among them. If

actors have a low *clique membership* 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 grouping together actors on the basis of similarity in which they are tied to.

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 in the number of *factions* we want. The output is a set of different groups which actors are more likely to be tied 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 GenialChef¹, a restaurant recommender system developed within the IRES Project². 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 composing this system can be grouped into service agents and personal agents (PA)(see the system architecture 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 assist PAs to find 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 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 agents as personal entities which you can 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

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.

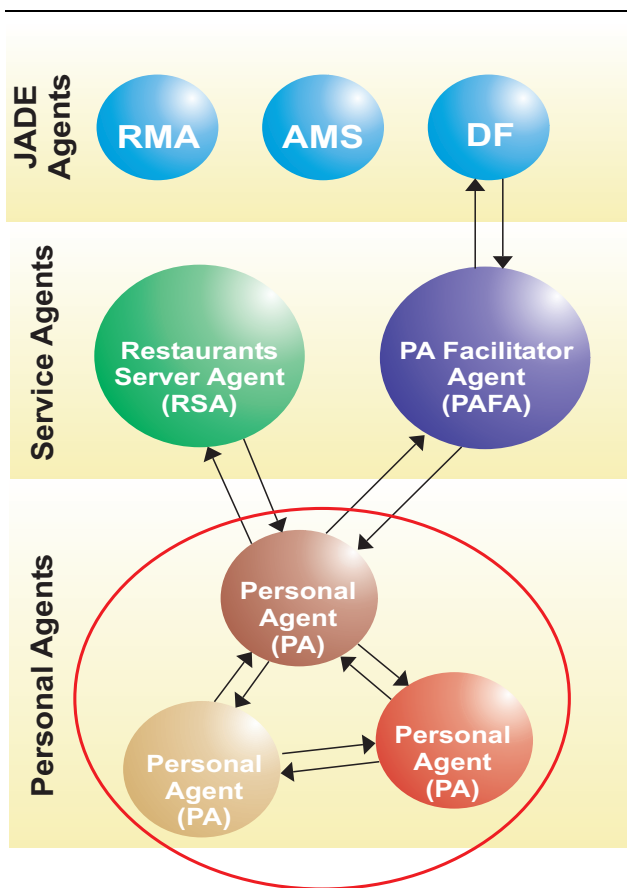


Figure 1. System Architecture

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, new items that could 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 as goes by depending on the results of the recommendations in order to improve acquaintance.

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

Thus, 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 out this relationships through the contact list of reliable agents each PA has. For example, Figure 2 shows the layout of a social network obtained in our experiments. What this figure shows

are relationships among PAs. Following the example, *sala_2* trusts in *pages*, *llado* and *mirocoll* opinions/advices and collaborates in *cunat_mangui* recommendations.

It is also important to have 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, A does not necessarily trust in C. Therefore, some of usual 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 4 to analyse how the collaboration is done in a concrete framework. In particular, the collaboration among PAs of the recommender system explained in section 2 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.

The simulator has some configurable parameters. The one which affects our analysis the most is the *trust threshold*. This parameter has two functions:

- In order to elaborate the contact list, the *trust threshold* is the minimum trust agent A must have in another agent B, so that B belongs to A's contact list.
- During the recommendation process, the *trust threshold* is the minimum trust an agent A must have in another agent B, so that A takes into account B's opinions and advices.

This parameter varies from 0 to 1, where 0 indicates an agent has no restrictions when he has to trust another and 1 indicates an agent is the most restrictive he can.

In this section we analyse a social network resulting from a simulation of 60 days long with 40 real user profiles extracted from our university staff. The simulation was performed with a trust threshold of 0.85 and with the optimal parameters studied in [9].

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

The first two measures to analyse are *size* and *density*. Since we used 40 user profiles to run the simulation, the *size* of the resultant network is 40. The maximum number of ties we could have in a fully saturated network of *size* 40 is 1560. In our network we only have 170 real ties, so the *density* of our social network is 10.9%. This measure is quite low and tells us that, in general, PAs do not have much

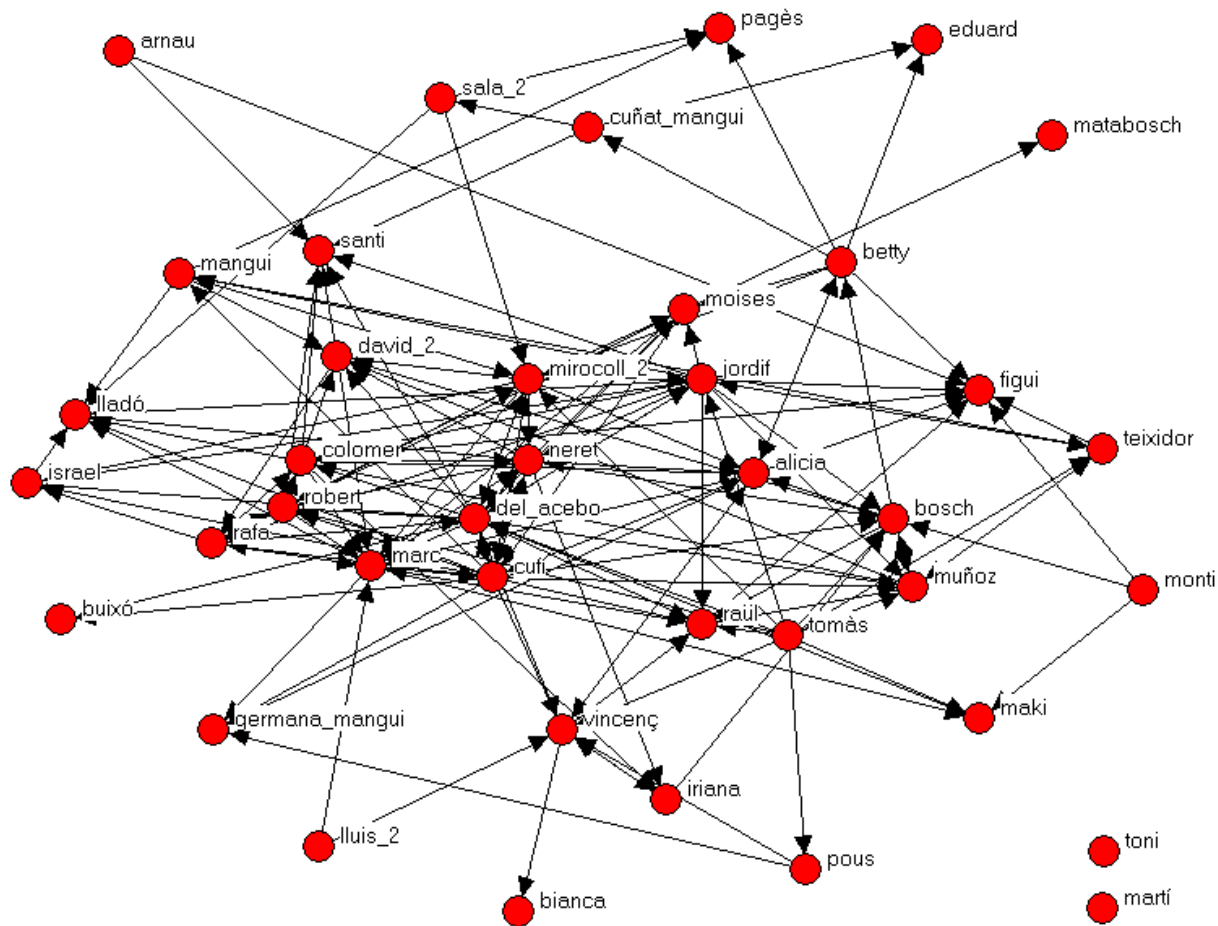


Figure 2. Graph layout of the relations created among PAs in our environment presented as a social network using Netdraw visualization program.

friends. We have used a high *trust threshold*, so it is difficult for each PA to establish a trust relationship with others.

Regarding *out* and *in-degree*, Figure 3 shows degrees for each PA. Higher *in-degrees* are those of *cufi*, *del_acebo* and *jordif*. These are the most prestigious PAs because there are more other PAs who trust them. The higher *out-degree* is for *marc*. Marc has 12 PAs in the contact list to ask for opinions/advices. If we look at Figure 2, we see all this PAs are drawn in a position which is quite central. Another thing to pay attention with is the fact that there is a large number (12) of PAs with an *out-degree* of 0. This means they do not have any PA in their contact lists and they cannot ask for advice in case of need. There are also 6 PAs with an *in-degree* of 0. This means they do not have any PA who trusts them. This is due to the fact that the pro-

files associated to the concrete PA are different from all others. In real world, there are people who like restaurants we would never go. If we ask for advice to people with different tastes than us and we get a bad recommendation, probably we do not trust them anymore.

Having analysed *degrees* for individual PAs, now we can analyse the *degrees* in the whole network. *Network centrality* for *out-degree* is 33.53% and for *in-degree* is 20.38%. We can say this values are quite low, and this result would indicate us that network is not much centralised. But we have to be aware of the fact that network density is really low. As there are very few ties in the network, this result may be altered. The mean value for *in* and *out-degree* is 4.25, and standard deviation is 4.61 for *out-degree* and 3.03 for *in-degree*. With this results, the *coefficient of variation* is 108.56% for *out-degree* and 71.32% for *in-degree*.

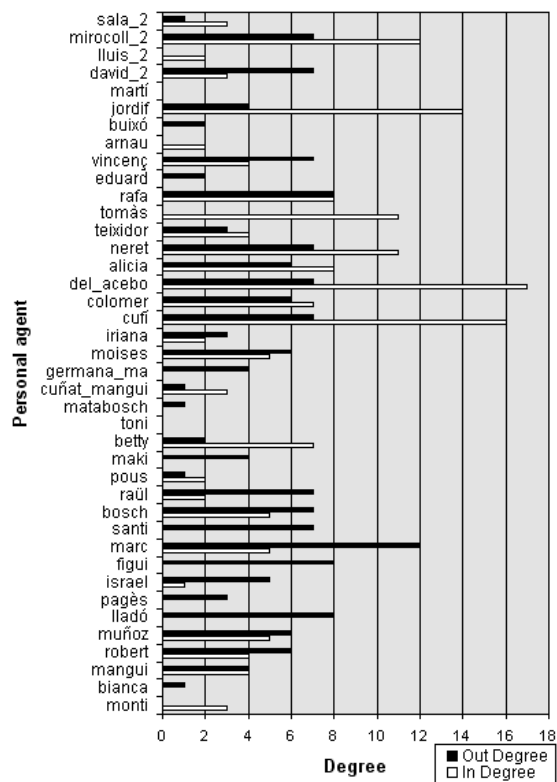


Figure 3. Freeman's Degree.

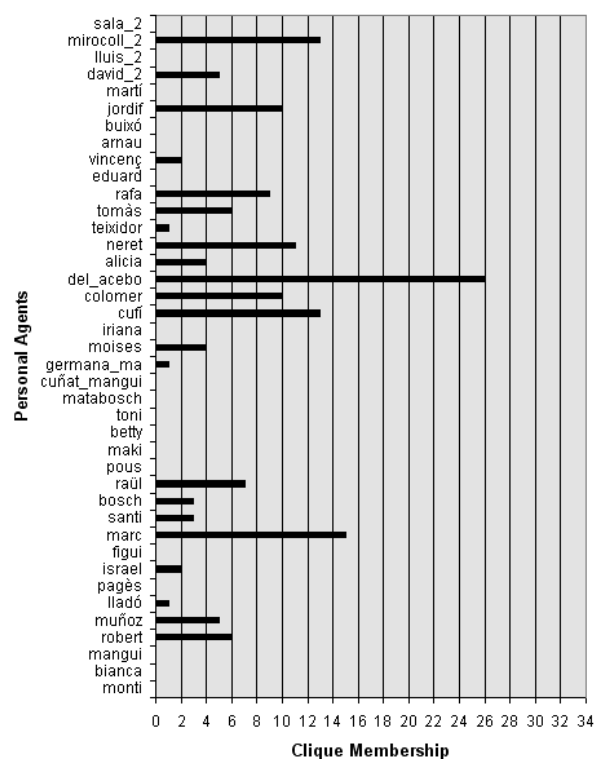


Figure 4. Personal agents membership to cliques of size 4 or more.

Clearly the population is more homogeneous with regard to *in-degree*, but the fact is that both values are high, so we can conclude that structural positions are heterogeneous and that the *network centralisation* is high.

Now we analyse possible network substructures. Figure 4 shows the number of *cliques* each actor is member. As there are several actors who have a high *clique membership* we can say with a high probability that exists at least one community in our network. There are 18 actors who have a 0 *clique membership*. This indicates that there are some actors who clearly do not belong to any community.

In order to corroborate the hypothesis of *clique membership* analysis, we analyse network substructures by means of *factions*. 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, vincenç, arnau,

buixó, germana.mangui, martí, toni, lluis.2, cunat.mangui, sala.2

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 5 to analyse them. In region 1-1 density is 0.28, in region 2-2 is 0.06, and in regions 1-2 and 2-1 densities are 0.05 and 0.04 respectively. On one side 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 them and as a consequence, they are the ones who collaborate the most. On the other side, *faction 2* has a low density. In fact, regions 1-2 and 2-1 have almost the same density than region 2-2 (*faction 2*). Therefore, *faction 2* cannot be considered as a community. This is the consequence of not having enough PAs in our 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 relation among the Internet

Thus, the next step in our work is to perform large-scale experiments with different parameters analysing in each of them the measures proposed. Taking into account the performance of the recommender system we will be able to see which the optimal model of social network that optimises the collaboration of the system is. In particular, we want to study how the best collaboration is obtained, for example, in a centralised/descentralised 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 particular, we want to study how relationships among users are generated, 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 *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.