Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 83 (2016) 496 – 503
Procedia Computer Science

The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016)

Self-Organizing Social Networks by Preference Similarity and the Networking Capacity of their Users

Debora Di Caprio[a,c], Francisco J. Santos-Arteaga[b,c,*], Madjid Tavana[d,e]

[a]*Department of Mathematics and Statistics, York University, Toronto, M3J 1P3, Canada*
[b]*School of Economics and Management, Free University of Bolzano, 39100, Bolzano, Italy*
[c]*Instituto Complutense de Estudios Internacionales, Universidad Complutense de Madrid, 28223 Pozuelo de Alarcón, Madrid, Spain*
[d]*Business Systems and Analytics Department, Distinguished Chair of Business Systems and Analytics, La Salle University, Philadelphia, PA 19141, USA*
[e]*Business Information Systems Department, University of Paderborn, D-33098 Paderborn, Germany*

Abstract

Consider the decision faced by the user of a social media site of whether or not to accept a friendship request from another user, given the limited amount of information available before deciding. We formalize the problem by defining the expected utility trade-offs derived from the request and simulate the resulting incentives numerically. These incentives provide the basis on which to build social networks determined by the different expectations and preferences of their users. Social networks are generated using a self-organizing map to cluster the decision makers (DMs) by their friendship acceptance behaviour. This behaviour is determined by the distribution of requesters relative to the preferences of the DMs.

Keywords: Social media; Expected utility; Preference similarity; Self-Organizing map; Neural networks.

1. Motivation

The emergence of social media has led to a substantial increase in the amount of personal information available about their users[1], leading other media users and companies to use this information strategically[2]. At the same time,

* Corresponding author. Tel.: +39-0471-013000; fax: +39-0471-013009.
E-mail address: fsantosarteaga@unibz.it

doi:10.1016/j.procs.2016.04.241

social media research generally focuses on identifying the main factors determining the structure of already existing networks, while acknowledging the existence of different types of users in terms of their networking capacities and influence on other users[3,4]. In this paper, we consider these characteristics of social media and their users but take a different research route from that of the existing literature.

We formalize the acceptance or rejection decision faced by a decision maker (DM) when receiving a friendship request and use the resulting framework as the base on which to build the corresponding network structures. In this regard, the friendship acceptance (or rejection) model defined in this paper relates to the basic postulates of expected-utility-based economic decision theory[5], where a DM makes a decision considering the highest expected utility attainable at a given point in time[6,7,8].

The DM has to decide whether to accept a given friendship request and generate a link expanding his network, or reject it and either find a more suitable requester or remain with his current set of friends. When receiving a given friendship request, some basic but important information becomes available to the DM, indicating the main preferences (i.e. likes, pages followed) of the person requesting his friendship. Once accepted, additional secondary information becomes available, which can be used by the DM to complete his profile of the requester. Consequently, we will assume that the initial information provided to the DM is correlated with the secondary one and, therefore, conditions its expected realization. Finally, the capacity of the requester to increase the network of friends of the DM must also be considered. This capacity should be determined by the connections of the requester and his similarity in preferences with the DM. As a result, the initial (observed) and secondary (expected) characteristics observed can be used by the DM to determine the expected networking capacity of the requester.

We formalize the above decision problem by defining the expected utility tradeoffs derived from the request and simulate the resulting incentives of the DM numerically, which, at the same time, provide the basis on which to build social networks determined by the different expectations and preferences of its users. Social networks are generated using a self-organizing map to cluster the DMs by their friendship acceptance behavior, which, at the same time, is determined by the distribution of requesters' characteristics relative to the preferences of the DMs. We illustrate how the differences between the subjective beliefs used by the DM to define his expectations and the distribution of characteristics across requesters condition the formation of clusters in the resulting network.

2. Basic assumptions

The choice made by the DM regarding the friendship request depends on the following variables:

- $X_1 = [x_1^m, x_1^M]$: The characteristics/preferences of the requester directly observable when receiving a friendship request. It accounts for publicly available information that describes the main basic tastes of (likes displayed by) the requester. The realization observed is related to the remaining information, which is unavailable at the moment of the request together with the list of friends and, therefore, the networking capacity of the requester.
- $X_2 = [x_2^m, x_2^M]$: The characteristics/ preferences of the requester that become observable after accepting the friendship request. It allows the DM to obtain additional information regarding both the tastes on the requester as well as his potential networking capacity. Thus, the distribution of this variable is related to and influenced by the realization of X_1 , while both X_1 and X_2 affect and determine the potential networking capacity of the requester consistent with the preferences of the DM.
- $X_3 = [0,1]$: This characteristic reflects the networking capacity of the requester. The shape of its associated probability function is determined by the realizations of both X_1 and X_2 . It should be noted that the friends of a given social media user can be classified in different categories, with access to different levels of information. However, even if not allowed to access the whole network, the DM becomes part of the group of friends of the requester. That is, even though the DM may not have the same status as other friends, who may be used to expand his network but are classified in a different category by the requester, he may still benefit from the fact that those potential network friends can actually observe him.

The acceptance decision of the DM will therefore be determined by two incentive functions defining the expected utility derived from either accepting a given friendship request or rejecting it. If the DM rejects the request, he must consider the probability of improving upon the current request in the future and compute the corresponding expected utility that would be derived. At the same time, in order for the DM to actually make a decision, both these functions must be determined by the values of all the potential realizations of X_1 that may be observed.

3. Accepting the request

As stated in the previous sections, the information available to the DM when deciding whether to accept the friendship request or reject it is limited to the initial observation of X_1. If the DM decides to accept the friendship request, his utility function, in expected terms, will be defined as follows

$$Accept = \int_0^1 \int_{x_2^*}^{x_2^M} B_3\left(x_1 + u_2^{-1}(E(x_2|\mu_2(x_2|x_1))), ce_1 + ce_2\right)\mu_2(x_2|x_1)u(x_1,x_2,x_3)dx_2dx_3 +$$
$$\int_0^1 \int_{x_2^m}^{x_2^*} [B_3\left(x_1 + u_2^{-1}(E(x_2|\mu_2(x_2|x_1))), ce_1 + ce_2\right)\mu_2(x_2|x_1)\left\{u(x_1,x_2,x_3) - c(x_1,x_2)\right\}]dx_2dx_3 \quad (1)$$

The following definitions and notations are required to interpret the above equation.

- The utility functions considered through the paper are given by $u_1(x_1) = x_1$, $u_2(x_2) = x_2$ and $u(x_1,x_2,x_3) = (x_1 + x_2)x_3$. Note that the first two characteristics are additively separable while the third one is used to generate the expected payoff obtained by the DM based on the networking capacities of his new potential friend.
- $\mu_i(x_i) = 1/(x_i^M - x_i^m)$ for $x_i \in [x_i^m, x_i^M]$, $i = 1,2$. The complete uncertainty of the DM regarding the distribution of potential friends within the population is reflected using uniform density functions, which are endowed with the maximum information entropy value.
- Following the standard economic theory of choice under uncertainty, we assume that the DM elicits the i-th certainty equivalent (CE) value induced by $\mu_i(x_i)$ and $u_i(x_i)$ as the reference point against which to compare both the observed and potential characteristics of a requester. Given $i = 1, 2$, the *certainty equivalent of* μ_i *and* u_i, denoted by ce_i, is a characteristic in X_i that the DM is indifferent to accept in place of the expected one to be obtained through μ_i and u_i. That is, for every $i = 1, 2$, $ce_i = u_i^{-1}(E_i)$, where E_i denotes the expected value of u_i. The continuity and strictly increasingness of u_i can be used to guarantee the existence and uniqueness of the $i-th$ CE value, respectively.
- A direct (subjective) correlation will be defined between the initial realization observed and the expected realization of the second characteristic of the requester. In this regard, the subjective probability defined by the DM on the set of potential realizations of the second characteristic should become an increasing function of the first characteristic observed. Given the uniform probability assumed on the first characteristic space, the density of the second characteristic will be defined as follows:

$$\mu_2(x_2 \mid x_1) = \frac{1}{x_2^M - x_2^m} + \varphi\left(\frac{x_1 - ce_1}{x_1^M - ce_1}\right)\frac{1}{x_2^M - x_2^m} \quad if\ x_2 > \frac{x_2^M + x_2^m}{2}$$
$$\mu_2(x_2 \mid x_1) = \frac{1}{x_2^M - x_2^m} - \varphi\left(\frac{x_1 - ce_1}{x_1^M - ce_1}\right)\frac{1}{x_2^M - x_2^m} \quad if\ x_2 \le \frac{x_2^M - x_2^m}{2} \quad (2)$$

with $\varphi \in [0,1]$ weighting the strength of the shift in the mass between the lower and the upper interval domains. Note that if the characteristic observed is equal to the CE value, there is no shift in mass between the intervals of the distribution. The above definition makes extensive use of the symmetry existing in the risk neutral (linear) case between the CE value and the extremes of the interval domain on which x_1 is defined. In this case, the CE value is located exactly in the middle of the domain. If this were not the case, the density function should be modified accordingly and the resulting shift in mass between both intervals adapted to the location of the CE value.

- Given the initial realization of x_1 obtained from the requester, denoted by x_1^o, the minimum value of x_2 required by the DM to derive an above-CE expected utility from the new friendship is given by x_2^*, with $x_2^* = ce_1 + ce_2 - x_1^o$. Whenever $x_2 < x_2^*$, the DM suffers a disutility of $c(x_1, x_2)$ from accepting the friendship of a requester whose tastes and characteristics differ significantly from his own. These disutility costs may include those derived from any undesired communication or further friendship requests from the network of the requester, together with potential negative effects on the friends from the network to which the DM already belongs.
- $B_3\left(x_3; x_1 + u_2^{-1}(E(x_2|\mu_2(x_2|x_1))), ce_1 + ce_2\right)$ is a Beta density function that will be used to represent the degree of optimism or pessimism of the DM regarding the networking capacity of the requester. The density function corresponds to that of the standard Beta distribution but its parameters are defined by:
  - The value of x_1 together with that of the secondary characteristic expected to be observed, which is determined by $\mu_2(x_2 | x_1)$.
  - The corresponding CEs, used as reference values for the evaluation performed on the information retrieved from the requester.

  Formally, we have the following definition of the Beta density for $0 \le x_3 \le 1$

$$B_3(x_3; x_1 + u_2^{-1}(E(x_2 | \mu_2(x_2 | x_1))), ce_1 + ce_2) = \frac{x_3^{x_1 + u_2^{-1}(E(x_2|\mu_2(x_2|x_1)))-1}(1-x_3)^{ce_1+ce_2-1}}{\int_0^1 u^{x_1+u_2^{-1}(E(x_2|\mu_2(x_2|x_1)))-1}(1-u)^{ce_1+ce_2-1}du} \tag{3}$$

where $u_2^{-1}(E(x_2 | \mu_2(x_2 | x_1))) = u_2^{-1}\left(\int_{x_2^m}^{x_2^M} \mu_2(x_2 | x_1) u_2(x_2) dx_2\right)$ is the expected realization obtained from the second characteristic of the requester given the one initially observed.

The potential capacity of the requester to network the DM with other similar users is based on both the initially observed and the secondary expected characteristic. Note that the shape of the functions $\mu_2(x_2 | x_1)$ and $B_3(x_3; x_1 + u_2^{-1}(E(x_2 | \mu_2(x_2|x_1))), ce_1 + ce_2)$ are determined by the initial realization of x_1. In this regard, Figure 1 illustrates the $B_3(x_3; x_1 + u_2^{-1}(E(x_2 | \mu_2(x_2|x_1))), ce_1 + ce_2)$ density functions considered by the DM for different realizations of x_1.

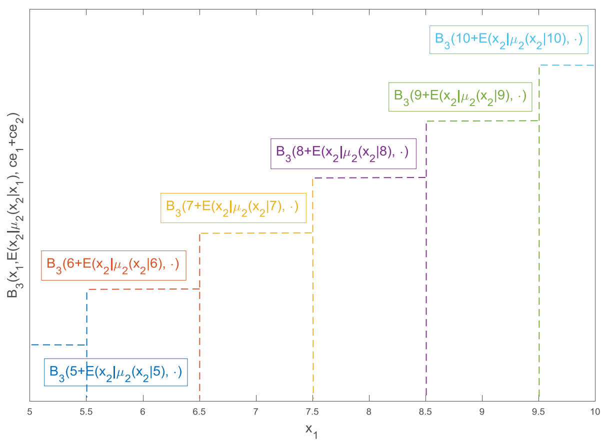


Fig. 1(a). Interval values of x_1 realizations associated with each $B_3(x_3; x_1 + u_2^{-1}(E(x_2 | \mu_2(x_2 | x_1))), ce_1 + ce_2)$ density

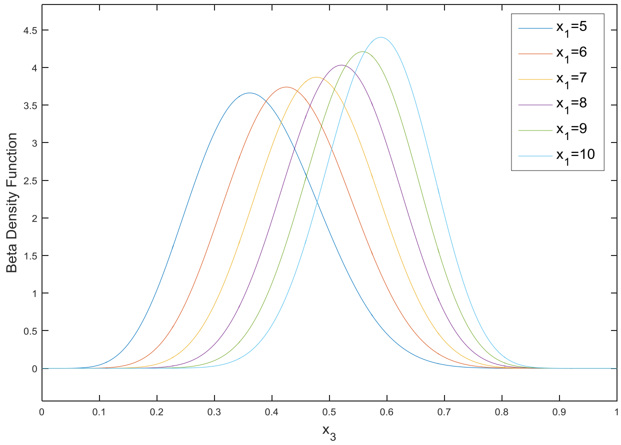


Fig. 1(b). $B_3(x_3; x_1 + u_2^{-1}(E(x_2 | \mu_2(x_2 | x_1))), ce_1 + ce_2)$ density functions for $x_1 = 5,6,7,8,9,10$.

Fig 1. $B_3(x_3; x_1 + u_2^{-1}(E(x_2 | \mu_2(x_2 | x_1))), ce_1 + ce_2)$ defined for the set of potential realizations of X_1 when $X_1 = [5,10]$ and $X_2 = [0,10]$.

In order to simplify the computations and account for the limited capacity of DMs to assimilate and manage information[9,10], we will consider different potential networking intervals defined by the DM in terms of the value of the initial characteristic observed. That is, similarly to the use of membership functions when defining a fuzzy variable over the interval domain of a set of potential alternatives, we have defined a set of Beta functions based on the potential realization intervals delimited within the domain of the first characteristic.

4. Rejecting the request

The rejection payoff is determined by the following expression

$$Reject = \int_0^1 \int_{x_2^*}^{x_2^M} \int_{x_1^o}^{x_1^M} [B_3\left(x_3; x_1 + u_2^{-1}(E(x_2|\mu_2\left(x_2|x_1\right))), ce_1 + ce_2\right)\mu_2\left(x_2|x_1\right)\mu_1\left(x_1\right)\left\{u\left(x_1, x_2, x_3\right) - s\left(x_1, x_2\right)\right\}]dx_1 dx_2 dx_3 +$$
$$\int_0^1 \int_{x_2^m}^{x_2^*} \int_{x_1^o}^{x_1^M} [B_3\left(x_3; x_1 + u_2^{-1}(E(x_2|\mu_2\left(x_2|x_1\right))), ce_1 + ce_2\right)\mu_2\left(x_2|x_1\right)\mu_1\left(x_1\right)\left\{u\left(x_1, x_2, x_3\right) - c\left(x_1, x_2\right) - s\left(x_1, x_2\right)\right\}]dx_1 dx_2 dx_3 -$$
$$\int_0^1 \int_{x_2^m}^{x_2^M} \int_{x_1^m}^{x_1^o} B_3\left(x_3; x_1 + u_2^{-1}(E(x_2|\mu_2\left(x_2|x_1\right))), ce_1 + ce_2\right)\mu_2\left(x_2|x_1\right)\mu_1\left(x_1\right) sc\left(x_1\right) dx_1 dx_2 dx_3 \quad (4)$$

- In the current setting, $s\left(x_1, x_2\right)$ denotes the search costs from observing the first characteristic of a new requester, which are incurred by the DM after rejecting the initial request.
- The first term of Equation (4) represents the expected payoff derived if the new requester provides a higher expected utility than the initial and the CE-based one. The expression is based on the realization of x_1 obtained from the previous requester, x_1^o, given the subjective distributions assigned by the DM to the X_1 and X_2 variables.
- The second term of the equation represents the expected payoff derived if the new requester provides an expected utility higher than the initial requester but lower than the CE-based one. As in the acceptance setting, $c\left(x_1, x_2\right)$

denotes the disutility from accepting the friendship of a requester whose tastes and characteristics differ from those of the DM. In this case, this disutility adds to the search costs incurred when observing the characteristics of a new requester.

- The last term accounts for the search costs $sc(x_1)$ incurred when the first characteristic of the new requester is lower than x_1^o. The new request is rejected and the costs incurred by the DM are assumed to account for the new search together with the disutility derived from not increasing his network of connections.

The accept and reject functions, together with the resulting acceptance and rejection intervals generated for $c(x_1,x_2)=s(x_1,x_2)=0$ and $sc(x_1)=1$, are represented in Figure 2.

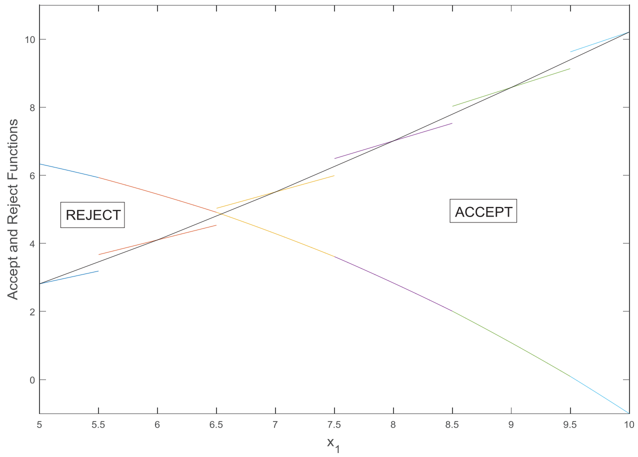


Fig. 2. Accept and reject functions with their corresponding intervals defined by $\boldsymbol{x}_1^* = 6.54$.

Note that we have used the x_1 reference value of each piecewise interval generated by the set of Beta functions to define a continuous approximation to the accept function. We should also emphasize that we will not analyze the effects derived from modifying the different search costs on the acceptance behavior of the DM, an analysis that can be considered among the potential extensions of the current model.

5. Clustering DMs through self-organizing maps

In order to generate a network structure where DMs are clustered into different groups by a self-organizing map (see Kohonen [11] and Sulkava et al. [12] for a detailed description of the main features of this type of neural network), we consider differences in the distribution of friendship requests in terms of the first characteristic observed by the DM. That is, differences in preferences are introduced between the requesters and the DM, who assumes a uniform distribution on the set of requesters due to his uncertainty regarding the distribution of potential friends within the population. We do so by defining two different Beta functions on the realizations of x_1 and accounting for the number of friendship acceptances when 25, 50, 75 and 100 randomly generated requests are received by 100 different DMs. The threshold value $\boldsymbol{x}_1^* = 6.54$ defined over the $[5,10]$ domain is transformed into $x_1^* = 0.308$ when considering the [0, 1] domain on which the Beta distribution is defined.

The results obtained after applying a self-organizing map algorithm to the number of friendship acceptances when the set of requesters follows a Beta$(x_1;2,4)$ and a Beta$(x_1;4,2)$ distribution are presented in Figures 3 and 4, respectively. Note that, given the location of the threshold value, the Beta$(x_1;2,4)$ scenario leads to a larger number of rejections than the Beta$(x_1;4,2)$ one. In both cases, we observe a central set of clusters surrounded by several isolated nodes. In the Beta$(x_1;2,4)$ scenario, these nodes may correspond to those DMs accepting either a larger or a lower number of friendship requests. Note that the dispersion exhibited by the position of the weights in this scenario is larger than the one obtained from the Beta$(x_1;4,2)$, whose isolated nodes correspond to those DMs accepting a lower number of friendship requests. Despite obtaining a larger number of lonely (isolated) DMs in this latter setting, we observe that they are more closely located to the remaining ones than those in the Beta$(x_1;2,4)$ scenario.

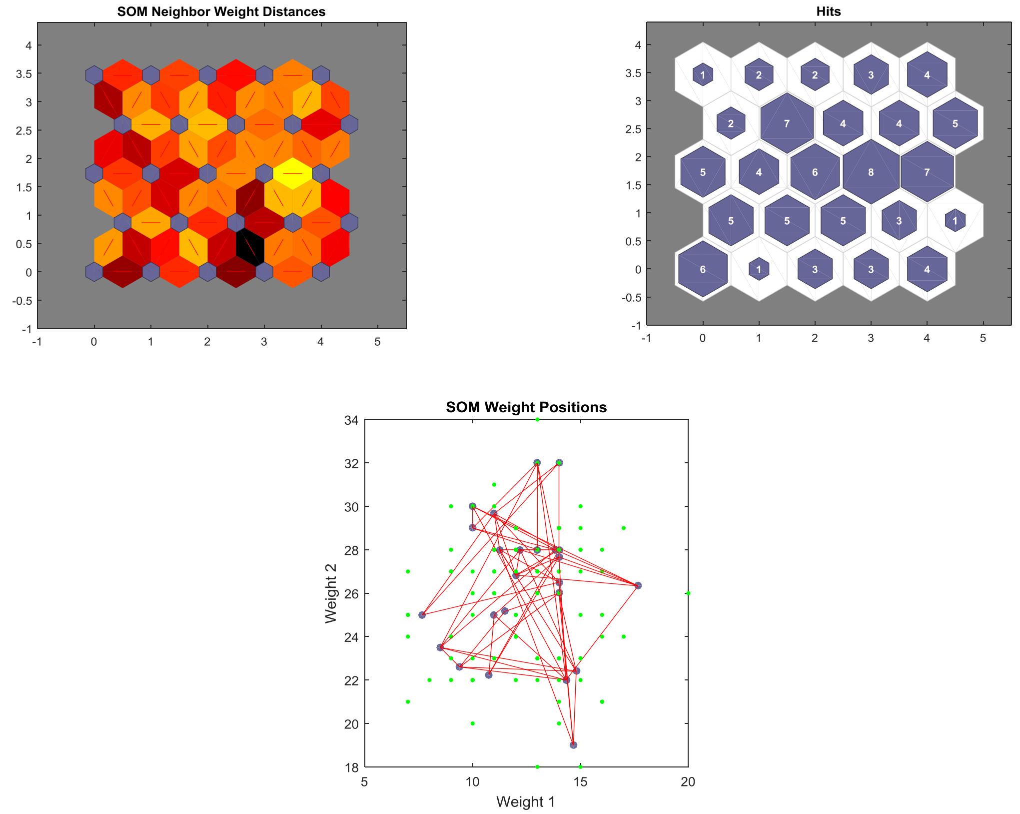
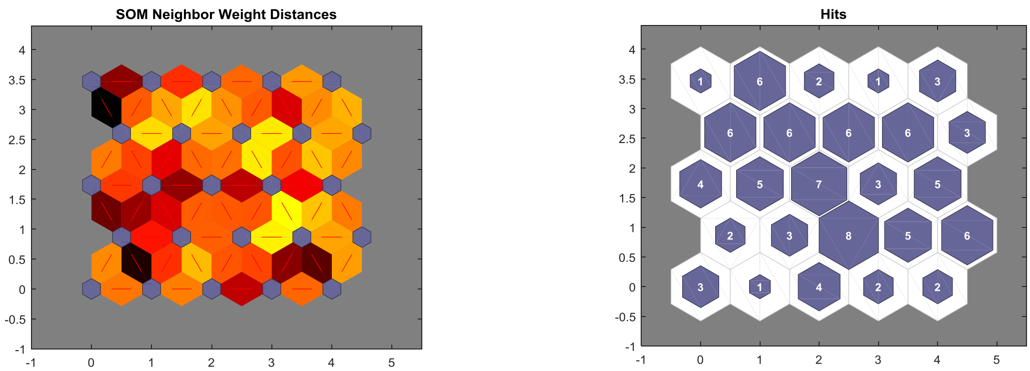


Fig. 3. Self-organizing map clusters following from a Beta$(x_1;2,4)$ distribution of requesters



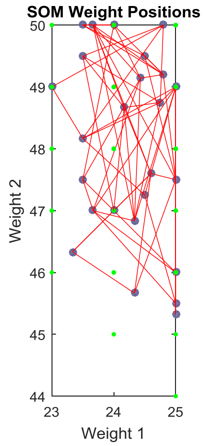


Fig. 4. Self-organizing map clusters following from a Beta $(x_1;4,2)$ distribution of requesters

6. Conclusions

We have developed a novel approach to analyze cluster formation within social networks. Our approach builds on the latest research on sequential information acquisition from a decision theoretical perspective[6,7,8]. This has allowed us to illustrate the type of cluster structures arising within a social network based on differences in the configuration of preferences among the users of a social medium. A self-organizing map algorithm has been implemented to cluster the DMs by their friendship acceptance behaviour.

We conclude by emphasizing that several variants of our decision model can be developed in order to consider, for example, multiple sequential observations within a finite set of friendship requests that may be expected to be received by the DM. This feature, together with variations in the degree of risk aversion of the DM, modifications in his subjective formation of beliefs, and differences in the distribution of friendship requests can be easily incorporated in the model and the clustering structures derived from the resulting networks analyzed.

References

1. Adamic LA, Adar E. Friends and neighbors on the web. *Social Networks* 2003; **25**: 211-230.
2. Stefanone MA, Hurley CM, Egnoto MJ, Covert JM. Information asymmetry and social exchange: exploring compliance gaining online. *Information, Communication & Society* 2015; 18: 376-389.
3. Guo H, Pathak P, Cheng HK. Estimating social influences from social networking sites - Articulated friendships versus communication interactions. *Decision Sciences* 2015; 46: 135-163.
4. Klein A, Ahlf H, Sharma V. Social activity and structural centrality in online social networks. *Telematics and Informatics* 2015; 32: 321-332.
5. Kahneman D, Tversky A. *Choices, values, and frames*. Cambridge University Press; 2000.
6. Di Caprio D, Santos-Arteaga FJ, Tavana M. The optimal sequential information acquisition structure: a rational utility-maximizing perspective. *Applied Mathematical Modelling* 2014; 38: 3419-3435.
7. Tavana M, Di Caprio D, Santos-Arteaga FJ, Tierney K. Modeling signal-based decisions in online search environments: a non-recursive forward-looking approach. *Information & Management* 2016; 53: 207-226.
8. Tavana M, Di Caprio D, Santos-Arteaga FJ. Modeling sequential information acquisition behavior in rational decision making. *Decision Sciences*. In press.
9. Simon HA. *Administrative behaviour*. Free Press; 1997.
10. Samiee S, Shimp TA, Sharma S. Brand origin recognition accuracy: its antecedents and consumers' cognitive limitations. *Journal of International Business Studies* 2005; 36: 379-397.
11. Kohonen T. *Self-organizing maps*. Springer Series in Information Sciences, vol. 30, 3rd ed. Berlin: Springer; 2001.
12. Sulkava M, Sepponen AM, Yli-Heikkilä M, Latukka A. Clustering of the self-organizing map reveals profiles of farm profitability and upscaling weights. *Neurocomputing* 2015; 147: 197-206.