

## SERENDIPITY IN SOCIAL NETWORKS

WERNER CREIXELL

Departamento de Electrónica  
Universidad Técnica Federico Santa María  
Av. España 1680, Valparaíso, Chile.  
Center for Spatial Information Science  
The University of Tokyo  
5-1-5 Kashiwanoha, Kashiwa-shi, Chiba, Japan

JUAN CARLOS LOSADA

Grupo de Sistemas Complejos, Departamento de Física,  
Universidad Politécnica de Madrid, 28040 Madrid, Spain

TOMÁS ARREDONDO

Departamento de Electrónica  
Universidad Técnica Federico Santa María  
Av. España 1680, Valparaíso, Chile.

PATRICIO OLIVARES

Universidad Técnica Federico Santa María  
Av. España 1680, Valparaíso, Chile.

ROSA MARÍA BENITO

Grupo de Sistemas Complejos, Departamento de Física,  
Universidad Politécnica de Madrid, 28040 Madrid, Spain

ABSTRACT. Serendipity is defined as fortunate discoveries made by chance. In this work we explore the idea that topological measures of a person's social network could be an indicator about how likely that person is to experience fortunate discoveries.

1. **Introduction.** Serendipity is the event of making fortunate discoveries by chance. Many use the term as synonymous with chance or luck. Serendipity is not a new concept and has been related to fortunate discoveries in many areas of inquiry and science. Recently, in the business administration area, company workers serendipity has been pointed out as one characteristic that would let companies stay competitive and ahead of the competitors [14]. Also there are claims that serendipity can be shaped in order to increase a person chances of making fortunate discoveries [9]. If these claims are true, they would imply that serendipity is not mere chance, and much more can be investigated about this phenomenon. However, to the authors best knowledge, there are no rigorous studies about this subject. We see this as a great opportunity to begin to investigate the phenomenon in a formal way.

---

2000 *Mathematics Subject Classification.* Primary: 58F15, 58F17; Secondary: 53C35.

*Key words and phrases.* Serendipity, social networks, complex networks, cellphone, clustering.  
The first author is supported by DGIP Grant 231021. He is also visiting researcher of CSIS at the University of Tokyo.

Our hypotheses are that serendipity is closely related with a person's social network and that serendipity can be measured in time as a function of observable variables in a person's social network. It is reasonable to assume, that the more active a person is in his/her social network, there are more possibilities for serendipitous discoveries. Additionally, it is also reasonable to assume that the more heterogeneous a person's social network is more diverse and useful the information, so that serendipity is more likely to occur.

In order to study the serendipity phenomenon, a data-set rich in users contextual data was used. The ideal platform from where extract such data is the cellphone, an ubiquitous device with increasing data processing, communicating and storing capabilities [8]. The cellphone contextual data was the one obtained in the Reality Mining project [3] at MIT. The data logged, by an application in the handset, included: call logs, SMS, nearby Bluetooth devices, cellular towers ID, application usage and phone status. Reality Mining data contains over 350000 hours of human behavior recorded during a nine month period. In that data-set, social network evolution in time can also be observed throughout this period of time.

Based on the ideas of serendipity previously described, the interactions among the users in the data set was studied using concepts from complex networks theory. Each individual in the data is made to correspond to a node, and the interaction between two persons was represented by a link between the corresponding nodes. In this work we have focused on two networks, one in which the links represent the Bluetooth encounters (figure 1a) and the other one created by the phone calls made between users (see figure 1b). In both networks we have searched for user behavior that could trigger serendipitous discoveries. Especially interesting were those nodes which had high connectivity and low clustering. This implies high social activity and heterogeneity for those particular nodes.

The interaction patterns and trends of the social network Twitter are identified in using longitudinal probabilistic social network analysis (SNA) [11]. An experiment with 128 Chinese students of English was performed. In particular, the relationship between the network structure in Twitter and the individual score of each participant in the project. The authors obtained three main observations from the experiments conducted: (i) A trend towards reciprocal communication, (ii) the number of communicating parties diminishes over time (selectivity) and (iii) students with higher scores received more attention from their pairs.

A profile classification of MySpace users is performed in [4] to discriminate commercial from individual users. Classification algorithms are used to determine which profiles are real users and which profiles are of enterprise or marketing users. For this, they used parameters like age distribution and usage patterns to make a decision tree classifier.

In [1], the authors detail the use of complex networks theoretical framework to characterize a vast number of different networks such as social networks, Internet, cellular networks, protein folding, etc. These ideas were also applied [10] to describe the behavior of potentials enemies in the context of fusion information (fusion information is the detection, correlation, combination and estimation of the information from multiple sources). Later, these concepts were used deeply in the same topic of fusion information in [5], using the Barabási-Albert model [1], to study models of fusion information.

The rest of this work is organized as follows: section 2 explains how the networks are constructed from the data, section 3 explores topological measures as serendipity

indicators, in section 4 a dynamic topological measure is proposed to single out users and finally in section 5 the conclusion and future work are discussed.

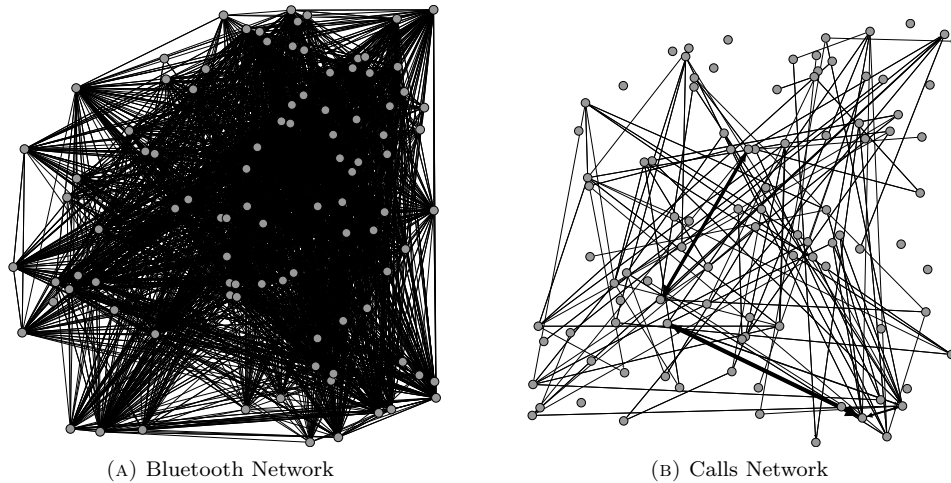


FIGURE 1. Representation of users interactions as weighted networks corresponding to Bluetooth network (left) and Calls network (right).

**2. Call and bluetooth networks.** The users interactions by means of phone calls and by Bluetooth (BT) encounters are displayed in the weighted networks of figure 1. Each node corresponds to a person in the reality mining project and the network edges correspond to the interaction between two users. Both networks are weighted, meaning that the weight associated to a particular edge corresponds to the number of interactions between the nodes of that edge. In the BT network the weights are the number of encounters between two people in the project, similarly the calls network consists of the number of calls between two people in the network but in this case the network is also directed. The calls network presents a much lower number of edges than the BT, and it is specially interesting because it only shows intentional interactions, while the BT network also includes incidental interactions.

Figure 2, shows the distributions of cumulative node degree ( $k$ ) and strength ( $s$ ). In the case of node strength, these are defined as the sum of the weights of their edges for the weighted network of calls and BT. It can be observed that 90% of the users have BT encounters with more than 28 different people ( $k_{BT}$ ) making a total of at least 87 encounters ( $s_{BT}$ ). Similarly, in the calls network 90% of the users call at least one person making or receiving two calls. On the other extreme of the distribution 10% of the users have BT encounters with more than 88 different people with a total of more that 5655 encounters. Similarly, 10% of the users calls (or receive calls from) at least to 7 different people with a total of more than 70 calls.

The clustering coefficient  $C^w$  for a weighted network is defined by equation 1, where the coefficients  $\{a_{ij}, a_{jh}, a_{ih}\}$  are elements of the matrix  $A \in \mathbb{R}^{n \times n}$  of adjacency, and  $s_i$  is node  $i$  strength, defined by the sum of all weights on edges from

TABLE 1. Average and maxima values for node connectivity degree ( $k$ ), clustering ( $C$ ), strength ( $s$ ) and weighted clustering ( $C^w$ ) for Bluetooth and calls networks.

Network	$N$	$\langle k \rangle$	$k_{max}$	$\langle C \rangle$	$\langle s \rangle$	$s_{max}$	$\langle C^w \rangle$
BT	97	64.082	95	0.819	2350.722	11115	0.848
Calls	97	2.227	11	0.170	77,175	893	0.181

node  $i$  [2].

$$C_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j \neq k} a_{ij} a_{jh} a_{ih} \frac{w_{ij} + w_{ih}}{2} \quad (1)$$

In the case of unweighted networks, the clustering  $C$  of node  $i$  is defined as

$$C_i = \frac{1}{k_i(k_i - 1)} \sum_{j \neq k} a_{ij} a_{jh} a_{ih} \quad (2)$$

Table 1 summarizes the maxima and average values for  $k$  and  $s$  and the average values for  $C$  and  $C^w$  for both networks.

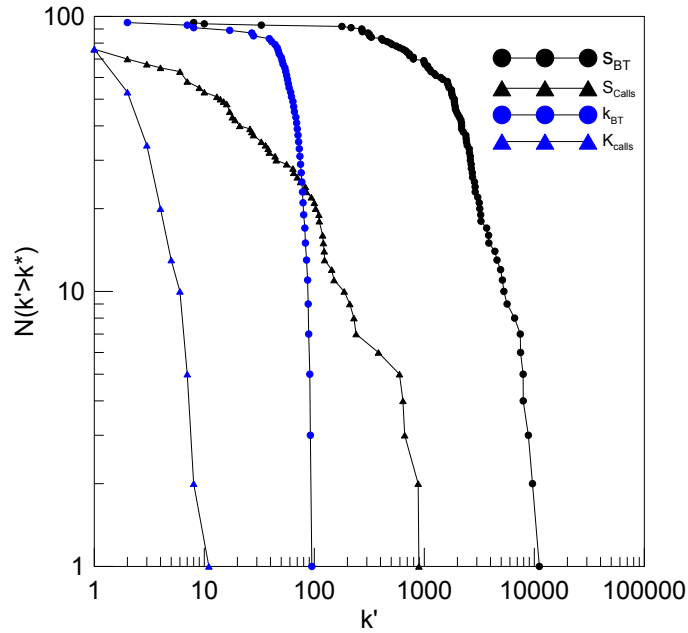


FIGURE 2. Cumulative node degree (blue) and strength (black) distributions for the Bluetooth (dots) and calls (triangles) networks.

**3. Topological network measures as serendipity indicators.** As previously mentioned the main assumption of this work is the idea that serendipity of a person is related to his/her social network. It is reasonable to assume that the person must be active in establishing new connections. Another important factor is that the person is able to establish connections with people from different backgrounds or social groups.

These two intuitive ideas can translate into two measures of network topology: total node connectivity degree,  $k$ , and clustering coefficient,  $C$ . The total node connectivity corresponds to the total number of people a person calls, in the case of the calls network, or encounters for the BT network. In simple words, a person with a high total  $k$  would mean a high degree activity while the clustering coefficient is a measure of how “related” are a person’s connections. We propose  $\zeta$  defined as

$$\zeta_i = \frac{k_i}{k_{max}}(1 - C_i) \quad (3)$$

as a intuitive metric for serendipity of node  $i$  for comparing persons of the same social network.

The measurements for both networks are displayed on figure 3. The relationship between  $k$  and  $C$  are plotted in figure 4 for the BT and calls network. The figure 3a) shows the relationship between the clustering coefficients in both networks. Similarly, in figure 3b) the relationship between the node degree  $k$  for all nodes in both networks is plotted. In figures 3c) and 3d) the same relationship for the corresponding weighted networks is presented.

In our present context, the interesting nodes are those for which the node degree is high and the clustering coefficient is low. The first condition means a person is active in his/her social network and the second condition means a person’s contacts are not very related among themselves. In figure 4 we can see that these conditions correspond to nodes located at the bottom right corner such as node 29, that present this behavior in both networks. In figure 3a) it can be noticed that the low clustering nodes are at the bottom left corner, node 29 is the node with significant activity that gets the closest to that region. Conversely, in figure 3b) we can see that the nodes on the upper right corner are those with a higher degree of activity in both networks. Accordingly, we can conclude the node 29 presents the most characteristic behavior for serendipity.

However it should be noticed that this behavior is not clear when we consider the weighted networks as can be observed in 3c) and 3d). The reason is that high values of the strength can be due either to be connected with many users or just to connect with few users but many times. In that sense, it appears that weighted networks do not seem to be useful in order to measure serendipity.

Node 85 is at the frontier between two different behaviors in the BT networks, as can be seen in the inset of figure 4. On the left side of node 85 (lower degrees) we can see that the nodes are dispersed in the graph, while on the right side (higher degrees) the nodes concentrate in a more clear trend. On the other hand, nodes 15 and 66, as seen in figure 4, are examples of nodes that present a tendency for higher clustering and lower activity than nodes 85 and 29 in both networks that were examined.

**4. Temporal evolution.** Serendipity is closely related to user behavior and it is reasonable to assume that it should change in time, as user behavior does. Therefore, we consider a temporally windowed serendipity measure ( $\zeta$ ). It is important

to observe the time evolution of users behavior in order to discriminate those that have consistent behavioral patterns in time. Figure 5 shows the time evolution for node  $i$ , in the BT network, in weeks ( $\Delta t = 1$  week), for the following function of network parameters:

$$\zeta_i^{\Delta t} = \frac{k_i^{\Delta t}}{k_{max}^{\Delta t}}(1 - C_i^{\Delta t}), \quad (4)$$

where  $k_i^{\Delta t}$ ,  $k_{max}^{\Delta t}$  and  $C_i^{\Delta t}$ , are calculated in the networks created by users in period  $\Delta t$ .

This user discriminant measure  $\zeta_i^{\Delta t} \in [0, 1]$  grows when both  $k$  and  $1 - C_i^{\Delta t}$  grow, therefore  $\zeta_i^{\Delta t}$  can be used to show the combined behavior of low clustering and high  $k$  for a particular user  $i$ . Therefore, when a user has a value of  $\zeta$  close to one means that he has high activity ( $k$ ) and low clustering (high  $1 - C_i^{\Delta t}$ ).

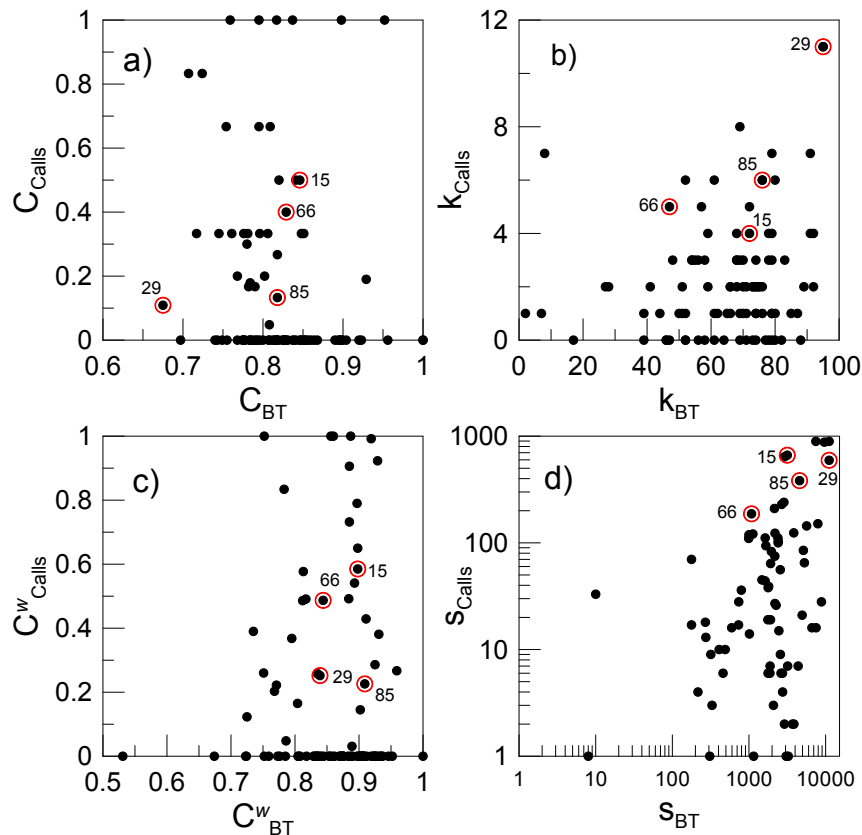


FIGURE 3. Comparison of the Bluetooth and call networks measures: clustering coefficients (a), connectivity degrees (b), clustering for the weighted networks (c) and strength degrees (d).

On the other hand, a  $\zeta_i^{\Delta t}$  value close to zero could be caused by low activity ( $k$ ) or high clustering coefficient. This means that even if the user has high activity but he mainly interacts with people within a fixed group, his clustering coefficient would be high and therefore  $\zeta_i^{\Delta t}$  would be low.

The weekly behavior, measured by  $\zeta_i^{\Delta t}$ , for users 29 (in red), 85 (in green), 15 (in blue), 66 (in yellow) and the average (in black) of all users is depicted in figure 5. We can see that most of the nodes have more activity between week 30 and 60 approximately. The value  $\zeta_{29}^{\Delta t}$  is the highest for all that period and it continues high even after week 60.

On the other hand, node 85, which is in the border of two different behavior patterns (as seen in figure 4), has a similar behavior as node 29 in the period from week 40 to 50, but before and after that it behaves close to the average.

Table 2 summarizes the network parameters and the values of  $\zeta$  of the users 15, 29, 66 and 85 for the global networks shown in figure 1 ( $\Delta t = 70$  weeks). It can be seen that for both networks, node 29 presents the highest value of  $\zeta$  (0.325 and 0.891 for the BT and calls networks respectively). It should be remarked that these values are almost double that the values corresponding to node 85.

**5. Conclusions and future work.** In this work we have analyzed the whole 9 month data of the reality mining experiment. The analysis was performed using the aggregated data for the whole period as well as weekly intervals of time. From aggregated as well as weekly data, we observed that node 29 behaves in a way that

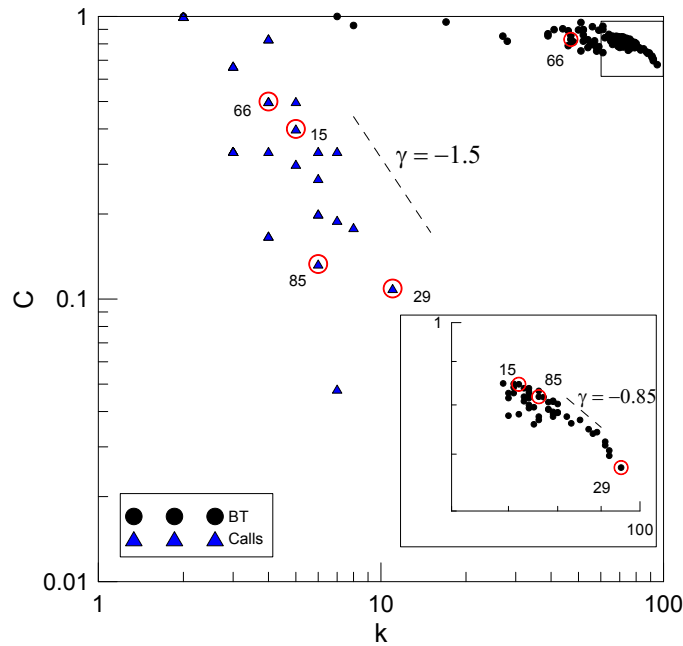


FIGURE 4. Clustering vs  $k$  for the nodes in the BT and calls networks.

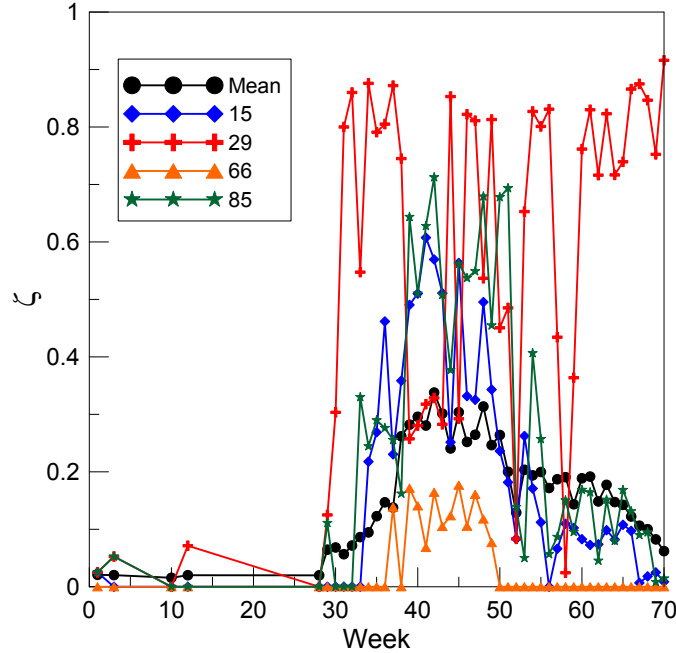


FIGURE 5. Measure of the function  $\zeta_i$  for users 29, 85, 15 and 66 in the BT networks.

TABLE 2. Values for node connectivity degree ( $k$ ), clustering ( $C$ ) and ( $\zeta$ ) function for both networks ( $\Delta t = 70$  weeks).

Node	$k_{BT}$	$C_{BT}$	$\zeta_{BT}$	$k_{calls}$	$C_{calls}$	$\zeta_{calls}$
Mean	64.082	0.819	0.1220	2.227	0.170	0.168
15	72	0.846	0.117	4	0.5	0.182
29	95	0.675	0.325	11	0.109	0.891
66	47	0.829	0.085	5	0.4	0.273
85	76	0.818	0.146	6	0.133	0.473

is more prone to establish new connections with people from different backgrounds, and it is reasonable to assume that from those new connections fortunate discoveries are more likely to occur than compared with the rest of the nodes in the network. We proposed  $\zeta$  as an indicator for this kind of behavior. It can be used to single out nodes with high  $k$  and low  $C$ . In this work we have proposed a hypothesis about serendipity and its relationship with the social network and behavior parameters (the Bluetooth encounters are more related to behavior), but there is lack of data in the reality mining experiment in order to fully verify our hypothesis. More experimental data are needed in order to verify if users like the user 29 are effectively having fortunate discoveries. As part of future work, it would be of interest to implement an experiment designed to generate a data set specifically tailored to



measure serendipity. This initial approach for studying serendipity only focuses in the number of interactions and not in the "quality" of those interactions, where quality could be the time duration of those interactions and frequency for example. Quality could also be interpreted as establishing a distinction between interactions, for example interacting with people whose  $\zeta$  value is high should be more valuable than interacting with people of low  $\zeta$ .

**Acknowledgments.** The authors thank the funding by the Universidad Técnica Federico Santa María DGIP Grants 231021 and 231138. Support from MICINN-Spain under contracts No.MTM2009-14621, and i-MATH CSD2006-32, is gratefully acknowledged. We also want to thank Nathan Eagle from MIT for kindly providing the Reality Mining Data-set and the associated documentation.

#### REFERENCES

- [1] Reka Albert and Albert-Laszlo Barabasi, *Statistical mechanics of complex networks*, Rev. Mod. Phys., **47** (2002), 47–97.
- [2] A. Barrat, M. Barthélemy, R. Pastor-Satorras and A. Vespignani, *The architecture of complex weighted networks*, PNAS, **101** (2004), 3747–3752.
- [3] N. Eagle and A. Pentland, *Reality mining: Sensing complex social systems*, Personal and Ubiquitous Computing, **10** (2006), 255–268.
- [4] W. Gauvin, C. Chen, Fu Xinwen and L. Benyuan, *Classification of commercial and personal profiles on MySpace*, IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), 21–25 March (2011), 276–281.
- [5] Feiyu Lian, Qing Li, Maixia Fu and Yuhan Zang, *A study on a general model of information fusion networks*, Pacific-Asia Workshop on Computational Intelligence and Industrial Application PACIIA'08, 2008, 579–584.
- [6] M. E. J. Newman, *The structure and functions of complex networks*, SIAM Review 2, **45** (2003), 167–256.
- [7] M. E. J Newman, A.-L. Barabási and D. Watts, "The Structure and Dynamics of Networks," Princeton University Press, 2006.
- [8] A. Pentland, N. Eagle and D. Lazer, *Inferring social networks structure using mobile phone data*, Proceedings of the National Academy of Sciences (PNAS), **106** (2009), 15274–15278.
- [9] Rawn Shah, "Serendipity Lies at the Heart of Business Relationships," Forbes November, 2010
- [10] Pontus Svenson, "Complex Networks and Social Network Analysis in Information Fusion," 9th International Conference on Information Fusion, (2006) 17pp.
- [11] K. Stepanyan, K. Borau and C. Ullrich, "A Social Network Analysis Perspective on Student Interaction within the Twitter Microblogging Environment," IEEE 10th International Conference on Advanced Learning Technologies (ICALT), (2010), 70–72.
- [12] Steven Strogatz, *Exploring complex networks*, Nature, **410** (2001), 268–276.
- [13] D. J. Watts and S. H. Strogatz, *Collective dynamics of small-world networks*, Nature, **393** (1998), 440–442.
- [14] In search of serendipity, Success in business increasingly depends on chance encounters, The Economist, July, 2010.

Received December 2011; revised June 2012.

*E-mail address:* [werner.creixell@usm.cl](mailto:werner.creixell@usm.cl)

*E-mail address:* [juancarlos.losada@upm.es](mailto:juancarlos.losada@upm.es)

*E-mail address:* [tomas.arredondo@usm.cl](mailto:tomas.arredondo@usm.cl)

*E-mail address:* [paolivaresr@gmail.com](mailto:paolivaresr@gmail.com)

*E-mail address:* [rosamaria.benito@upm.es](mailto:rosamaria.benito@upm.es)