"Website Digg" Project

Soraia Felício

Analysis of Social Networks and Information
Professor: Carlos Soares

Agenda

1) Context of the project

2) Problem definition

3) Results general

4) Results in relation to a particular user

5) Conclusions

Context of the Project

Digg is a news aggregator that includes linking or sharing stories around a particular topic.

Data analysed in this Project is about communication: each node in the network is a user of the website, and each directed edge denotes that a user replied to another user.
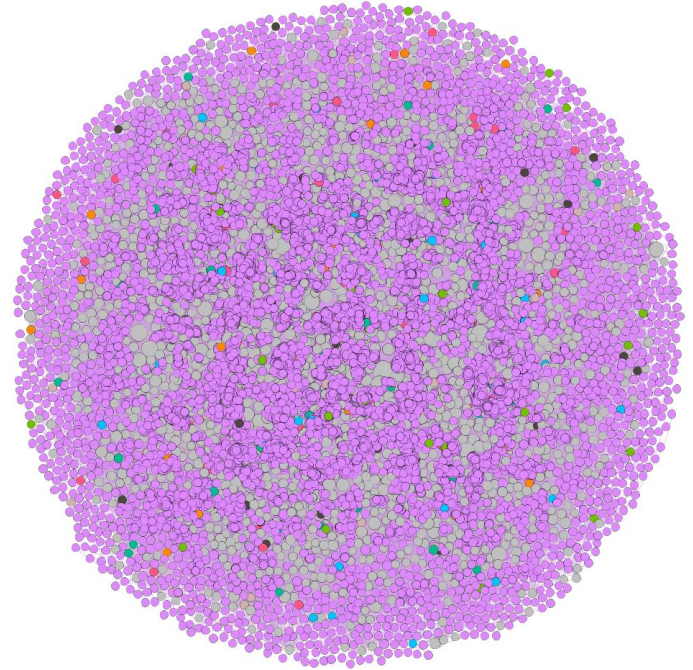
Context of the Project: Information of the Digg network

Type Information	Digg network
Repository	Konect - Koblenz network collection. Has **three categories of the Digg:** communication, rating and social.
Format	Directed
Nodes (users)	30,398
Edges (replies)	87,627
Category	Communication
Tools	Gephi and Jupyter

Problem definition

In a more general context, how to evaluate in a network "who-talks-to-whom" the groups that are formed in a certain period of time and what conclusions we can obtain with these analyzes, if we disconsider the context.

This analysis was performed in the part of the social network Digg that has only one user action (which each user responded to other users), in a given time interval.

1) Is it possible to find that seed user when we disconsider the context?
2) How to rate the most popular user on a network "who-talks-to-whom"?
3) Can we predict some synchronism of behavior in this network from the perspective only in the action of responding to the other user without considering the context?

Results general

Statistics of the Gephi in relation to **nodes (users)**	Used to...	Result	Meanings for the project
Clustering coefficient	probability to form triangles.	0.008 in relation to average	There is the possibility that at least one set of users are friends.
Modularity class	it divide the nodes in group or class	0.014 Modularity Coefficient 377 communities	Regarding the small world and the Watts-Strogatz model, the large amount of communication reveals that although a significant number of users on this network are not neighbors, they can communicate in a few steps until reaching the desired user.

Results general

Statistics of the Gephi in relation to nodes (users)	Used to...	Result	Meanings for the project
Strongly Connected ID	strong and weak ties	23,652 strong connections and 373 weak connections	This result reveals that network users are strongly connected and the few weak ties that exist allow the formation of local bridges. What confirms the small world model, through a few users it is possible to reach the most distant user.

Results general

Statistics of the Gephi in relation to **nodes (users)**	Used to...	Result	Meanings for the project
In-degree	** popularity => in-degree gregariousness* => out-degree * **the quality of enjoying the company of other people**.(Cambridge Dictionary)	73.23% of the nodes in-degree less or equal than 2	It is necessary to evaluate these users more deeply to know what type of influence they are exercising in the entire network, for this it is necessary to correlate with the topics they are interacting with.

** Source: Needham, M. and Hodler, A. E. Graph Algorithms: Practical Examples in Apache Spark and Neo4j 'Reilly Media, p. 79, 2019

Results general

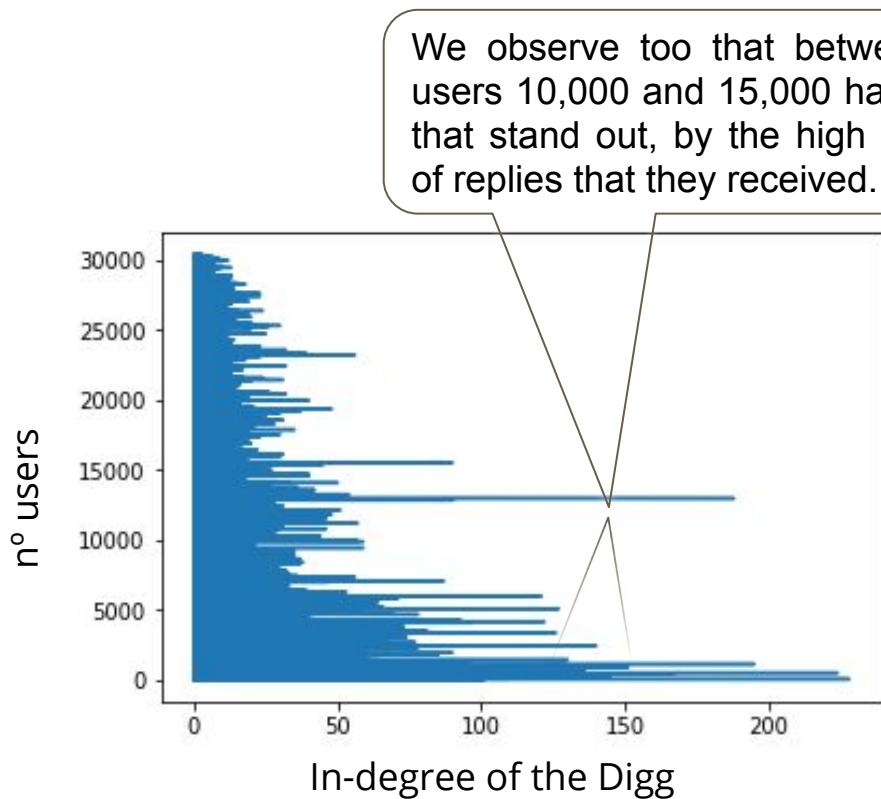| Filter of the Gephi in relation to **edges** (communication)	Used to...	Result	Meanings for the project
Mutual edges	To evaluate the edges of both direction. To evaluate the positive and negative relationship	154 nodes and 184 edges	This number of nodes and edges are considered a positive relationship, ie, there is an exchange of messages that suggests an interaction and a relationship. This amount also reveals that the number of interactions is low in relation to the total number of users on the network.

Results general: Homophily

Statistics Information of the Koblenz	Values	Meanings for the project
Assortativity	0.0046449	Low ability to associate with other users. Suggesting that there are few similarities within the network.
Reciprocity	1.55%	That indicates little interaction.
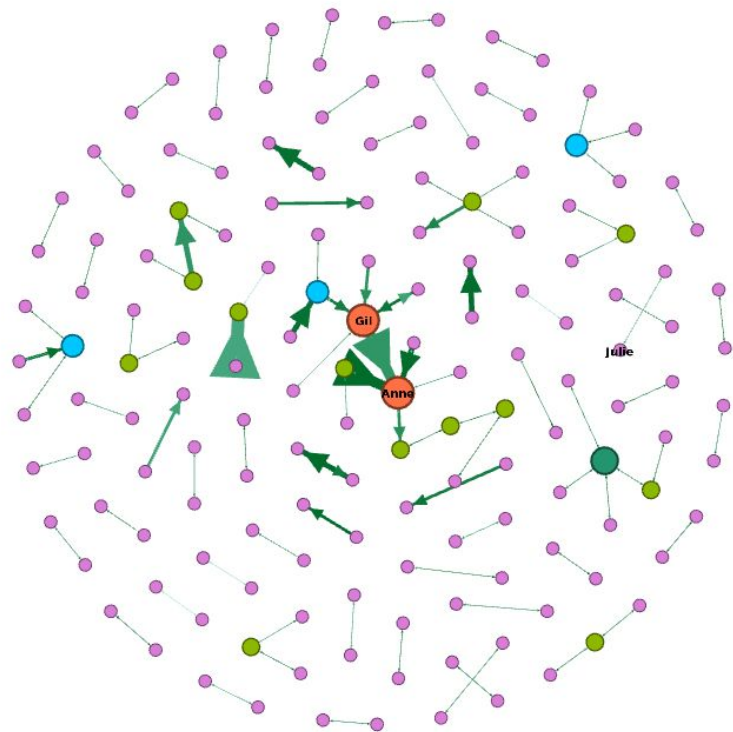
- There is a set of users that could be friends or else the users can be divided in two groups, in that at least a pair in each group like each other, and at least at the other end the pairs are enemies.

- This is observed with the relationships that we consider to be positive, because there is interaction between them and the relationships that have not interactions between them, we consider negatives.

- The graph on the side shows only the users who interacted. We noticed that there are few users that have interactions between them.

- Addition on that the users with fictitious names: Gil and Anne stood out in relation to the interactions. The user Julie had not very prominent, but she will be referenced at another time when we evaluate the popularity and that is why it was presented in this graph.

Results general: Long tail

We observe too that between the users 10,000 and 15,000 has some that stand out, by the high number of replies that they received.
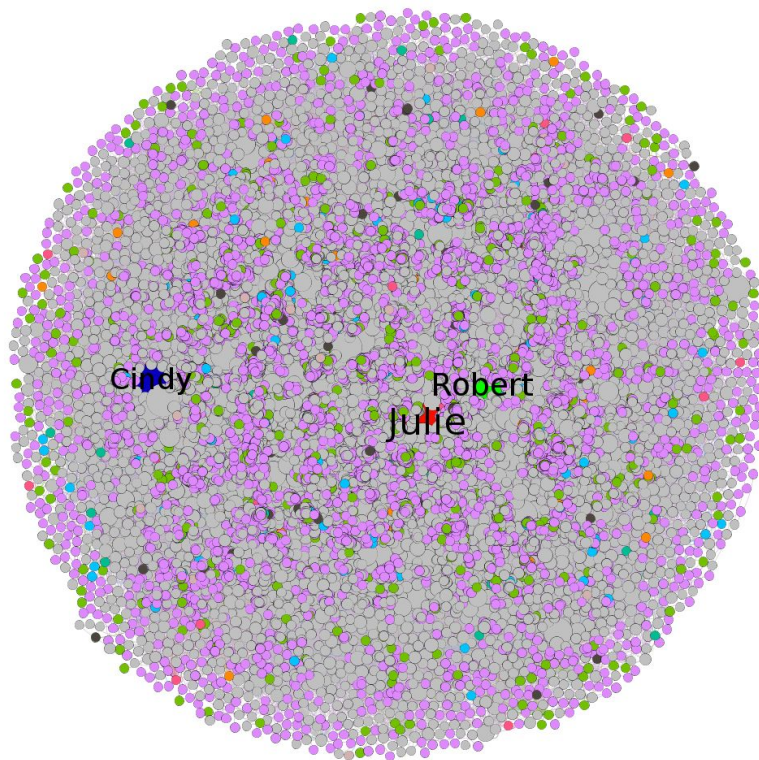

n° users

In-degree of the Digg

This concept is based on the principle that popularity is only for a few, we see from the graph on the side that. In relation to Dig, popularity is concentrated among the first 15,000 users. Justifying the choice of one of the most popular users, who is among those 15,000 users.

Results general: Rich-get-Richer in relation to Digg

Id	Label	Eigenvector Centrality ∧
72	Julie	1.0
511	Robert	0.789051
476	Cindy	0.73003

This network has some user more popularities, for example, Julie, and Robert and Cindy. This situation is observed by neighbors quantity (eigenvector centrality metric) or by degree of these users.

Some test with particular user and more results

To answer the research questions, some tests were done on a specific user. Following are more tests with three users who are related between them, through their replies in the Digg.

Results correlated with Cascading Behaviour

This user receives many messages suggesting that his interests arouse interest in other users. Somehow, his behavior can influence others.
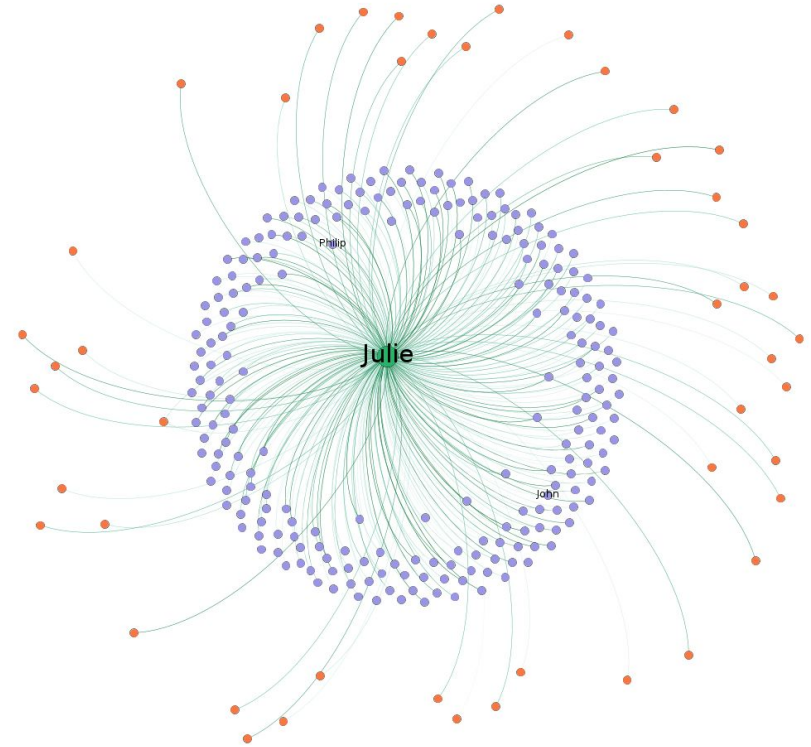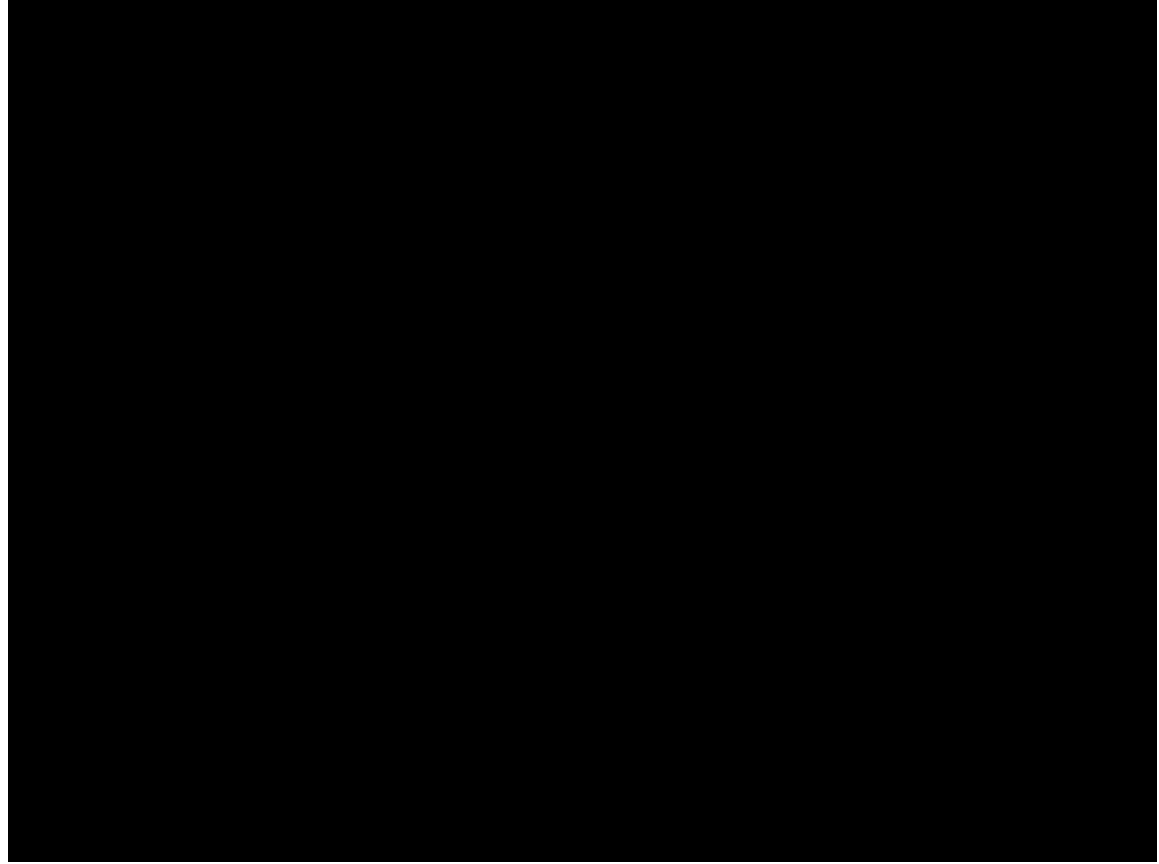
Results correlated with the Small-World

- This is proven when choosing other users close to him who form a local bridge in relation to the communities they form. Suggesting that the behavior of the previously chosen user may influence other users and communities.
- This is related to the concepts of homophily, behavior cascade and epidemic model.
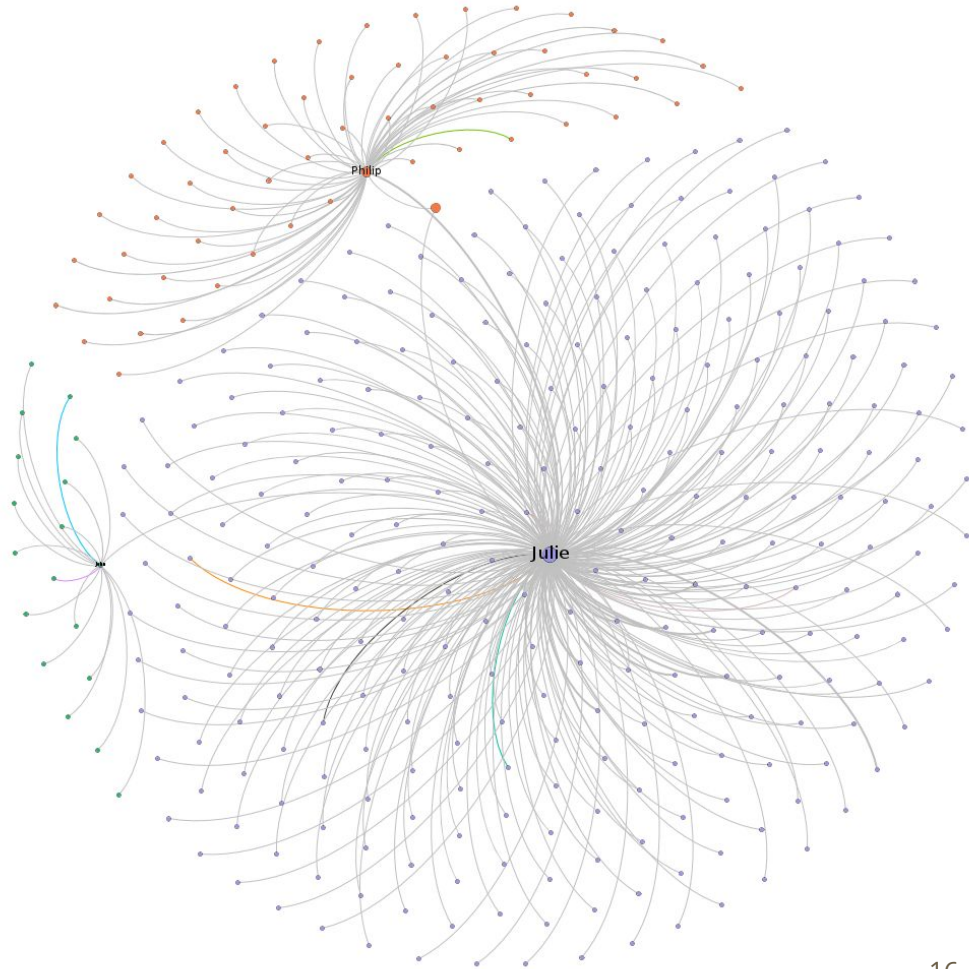
Results correlated with the Epidemic model

t1   t2   t3

Results correlated with the Game Theory

- If there are an action to reply or not then it is 1
- If there are not action 0

- Each box represents (Julie, Philip)

		Philip	
		Reply	**No Reply**
Julie	**Reply**	0,1	0,1
	No Reply	1,0	1,0

Julie doesn't reply to Philip, but Philip replies to Julie. We see that Julie is once again influencing Philip of some form. Thus, stimulating interaction between the users is not a Digg strategy, but it could be.

Discussion

These analyzes reveal some clues about the research questions raised and previously and about the network in general:

1) The network does not have much interaction, although it is strongly connected, but this is justified because it is a news classification network as main focus and not interaction between users. This is perceived by the large amount of communication that involves the network as a whole, but with few communications carried out in both directions (positive relationship). In addition to some evidence of similarities, that is, homophily.

2) There are a few users who stand out in relation to the number of responses they have, justifying the existence of a great number of users strongly connected, because this users of any way influence others users, arousing more connections between them.

3) We consider seed users, the users have high degrees and eigenvector centrality, ie, a lot of popularity.

4) The most popular user suggests that there is a synchronism and influence in relation to other users. As it is a small world it can spread its influence and impact the social network as a whole. Digg could consider in its functionalities topics that these users show more interest or the functionalities of following them, as it exists in other social networks.

Conclusions

- We identify the seed user, using the metric that measures popularity (in-degree and eigenvector centrality), thereby showing how to identify the popularity of a "who-talks-to-whom" social network.

- We showed a social synchronism between three users over a period of time without considering the context. Our conclusion is that there are few popular users, as described by the concept "long tail", but they are able to influence the behavior of several other users of the social network.

- We recommend that the interests of these users are observed so that in a network like Digg it will create new features, such as how to follow more popular users or recommend the same news of interest to these users, in addition to the news already recommended directly by users of the network by vote.

- As a future work, we also recommend looking at users who have many interactions to assess whether these users form groups or whether they are just part of a group and do not have much influence within their groups or are simply people who like to relate. Thus, how can a social network benefit from this type of person and help other people.

References

Data source: http://konect.uni-koblenz.de/networks/munmun_digg_reply

Easley, D, and Kleinberg, J. **Networks, Crowds, and Markets: Reasoning about a Highly Connected World**. Cambridge University Press, 2010.

Munmun De Choudhury, Hari Sundaram, Ajita John, and Dorée Duncan Seligmann. **Social synchrony: Predicting mimicry of user actions in online social media.** In *Proc. Int. Conf. on Computational Science and Engineering*, pages 151--158, 2009.

Needham, M. and Hodler, A. E. **Graph Algorithms: Practical Examples in Apache Spark and Neo4j** 'Reilly Media, 2019.

Thank you!