EXPANDED SOCIAL CIRCLES EFFICIENT QUERY ON SOCIAL GRAPHS

Rui Manuel Lima
Universidade Lusíada de Vila Nova de Famalicão, Portugal

ABSTRACT

Social graphs can be used to express personal relationships between internet users. Query the entire network is a very expensive task and here is no interest to continue a query when a successfully answer is already known. This paper proposes a searching mechanism to query nodes in several social graphs topologies. Active users are continuously expanding the social network, resulting in frequent changes on the social graph topology. The results show that expanded social circles are an efficient query mechanism and resilient to distinct social graph models.

Keywords. Searching mechanisms, online social networks, knowledge searching.

1. Introduction

On online social networks (OSNs) such as Facebook, Twitter and LinkedIn the relationships (ex: friendship) between users or organizations can be represented by undirected graph networks. The users are nodes and the relationships are the edges. The strength of relationship between users is out of this scope, but there are several models to deal with strong ties (Xiang, Neville, & Rogati, 2010). Graph models are interesting alternatives to datasets, because it is possible to generate multiple synthetic graphs that are statistical similar. The random graph models has been applied to study several networks (communications, transportations, metabolic) and social networks (Sala, et al., 2010).

Social networks can have access to sensor data that is gathered by many deployed applications such as: environmental monitoring, logistics, management and intelligent spaces, enabling pervasive computing between users and sensors. The sensor information of mobile devices is available to be shared on social networks (Lane, et al., 2010). For example: The temperature acquired by my brother's phone can be very important if we share the same physical space. Such simple query is an example of what can be easily done by smart-phones, although it raises several privacy issues. The user must be able to decide if it will share his private data, and what kind of information will be available. Several experiences are exploring these ideas, such as CenceMe (Miluzzo, Lane, Eisenman, & Campbell, 2007), where users from the same mobile social net- work can sense information that is important to them, but produced by someone else.

Each user action/decision is autonomous and independent from other user actions. In this research it was assumed that no central server has the knowledge to infer future user actions. However, several classes of queries can take advantage of the social network graph, to select which users should be addressed by the query. One example is the neighbour query, which search for neighbour nodes (friends) and spread the query to them. These relationships can be enhanced with privacy tools, which can classify social circles and sub-cirques according to users trust and intimacy. Rating data can be combined with OSNs information to infer circles of friends, which can be different from the explicit circles (Yang, Steck, & Liu, 2012).

But, what to do when the query didn't get the answer from the user social circle? One strategy is to expect that someone (that are also interested in the query) will autonomy decide to forward the query to other social circles, expanding the query scope to others. With this mechanism queries can be propagated in social networks according to the user decision, avoiding to be blinded delivered to everyone. Query expansion techniques can be used to get diversified

results (Radlinski & Dumais, 2006). Finding different users opinions, must take in account that querying all nodes (flooding) available in OSNs will consume too much time. This problem can take a big dimension and is practical infeasible, because network graphs are too large to fit on a single server (Sarwat, Elnikety, He, & Kliot, 2012). Also it must be ensured that the query process is stopped when the answer is found.

We propose a new query mechanism **Expanded Social Circles (ESC)**, where the user action is mandatory for query propagation, avoiding flooding the entire social network, and giving feedback to all the users that were responsible in forwarding the query to its social circles.

The paper is organized as follows. First a discussion of related work in Section 2. In Section 3 the proposed query mechanism is detailed presented. Then a description of the evaluation using graph models and simulation techniques on Section 4. Finally, provide the main conclusions and future work in Section 5.

2. Related Work

Nowadays users from social networks can create friends circles to share data with close friends, just like in real life. To enhance these facilities, we are proposing to give the users the possibility to forward queries that are requested from a friend, enabling the query to expand outside the initial social circle. Access control policies can be used to limit the queries of private data (Bjorklund, Gotz, & Gehrke, 2010), and those queries will not expand. There are some strategies to find the correct user to answer the query, using social network structure, semantics, terminology and tagging (Duchateau, 2011), but in this research no previous data is known, and all users have the same probability to give a successfully answer.

The query problem has some similarities with the first phase of searching resources on mobile ad-hoc networks (MANETs). In MANETs nodes can delay the queries in a multi-hop fashion to implement more energy efficient searching mechanisms. The Expanding Ring Search (ERS) (Park & Pu, 2007) is a query mechanism that avoids flooding the entire network, i.e. queries first the neighbours nodes, and if no answer is found then initiate a new search with a large ring to expand the query to other people. The search ends when the query initiator receives a successfully answer. The Broadcast Cancellation Initiated on Resource (BCIR) (Lima, Baquero, & Miranda, 2012), considers a multi-hop wireless network where the search process is initiated by node n_0, and n_H is the node (or one of the nodes) that has the nearest resource copy, and is located at H hops from n_0. The search process is initiated in node n_0 with the transmission of a search/query message m_s, to be disseminated using the intermediate nodes to forward (relay) the packets. When n_H receives the message m_s, a point-to-point (following the reverse route that is piggybacked to m_s transmissions) answer message ma is sent from n_H to n_0, to notify the discovery. The search process assumes that any node at any instant can start the diffusion of a cancellation message mc to stop the propagation of m_s. The above mechanisms were the inspiration to new efficient query mechanism for OSNs, that we are proposing. The algorithm details are presented in the next section.

3. Searching by Expanded Social Circles

In the expanding mechanism, illustrated in Fig.1, the query starts at n_0. Them, in Fig.1(a) are represented the n_0 closest friends, corresponding to the potential first hop of nodes. After nodes from hop 1 forward the query, a new subset of nodes is reached, as depicted in Fig.1(b), that correspond to the second iteration hop.

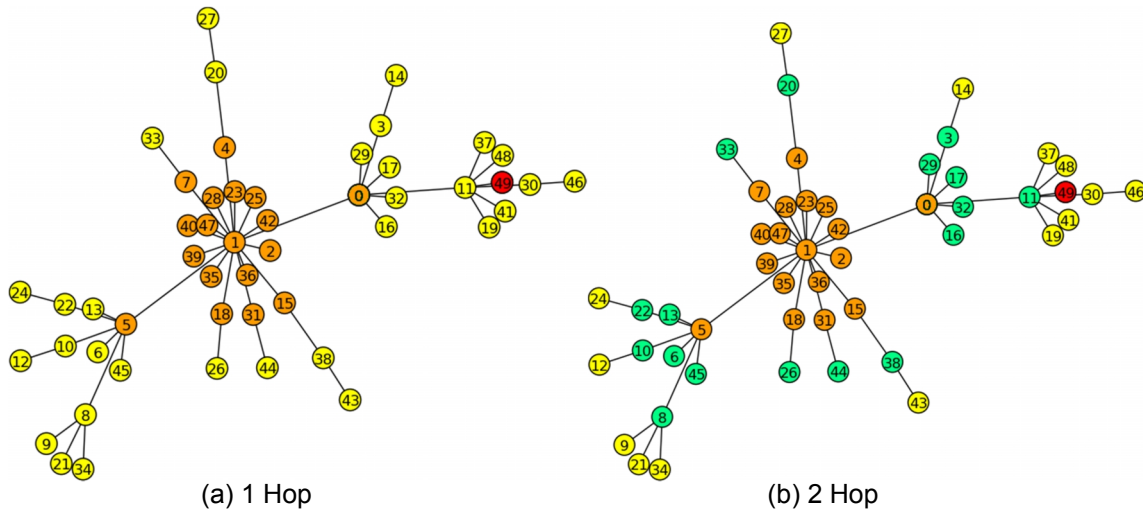(a) 1 Hop   (b) 2 Hop

Fig.1. Searching by Expanded Social Circles

In contrast to flooding the entire network, we propose a new mechanism **Expanded Social Circles (ESC)**, where the searching process is initiated by the node n_0, and n_H is the node (or one of the nodes) that knows the answer to the query. The hop H is the number of intermediate users that are in the path between n_0 and n_H. The search process is initiated in node n_0, sending the query message m_s to its neighbours. After receiving m_s, each node will wait for the user action: answer, forward or cancel the query. The process will end if the correct answer is found or the query is cancelled. If the user decides to forward, the query will be disseminated using the intermediate nodes. When n_H receives the message m_s, an answer is sent to n_0, directed by the central server and to all forward nodes that were interested in the query, as depicted by Fig.1. Waiting for the user action can increase the delay to get a good answer and the latency of the searching process is mainly influenced by the user response time.

The used system model is in-line with the related work, where nodes transmit in synchronized rounds and form network topologies. The social network is modeled by the graph G. Each graph G = (V, E) is defined by a set of vertices V (which represent the users) and a set of links between the vertices E (representing the friendship circle).

The overhead is the main metric used in evaluation, and is the ratio between the number of forward nodes until the search process ends and the number of network nodes. The user stress is quantified using the overhead metric, and it will consider all the needed interactions until the search is complete. Too much overhead has been considered as one of the reasons for turning away from social networks, when users are too busy and have no time to maintain what they considered an appropriate level of network interaction.

4. Evaluation

The evaluation explores the impact of network topology changes in the ESC mechanism and compares it to the flooding mechanism. Large graphs resulting from social networks can be very demanding when kept in memory for each query, but some solutions (Crecelius & Schenkel, 2012) explore weight graphs to compute the nearest neighbours (friends). This evaluation will not replicate the size of the online network, it will assume only a small sub-network connected component. During the evaluation tests, the simulator generates thousands of random network graphs (for each model), to extract the average characteristics of the query process for each topology.

The evaluation focus on finding resources that are in the connected component of the querying node, and, without loss of generality, considers that the network has a single connected component. This analysis assumes no nodes failures. For each search operation the simulator generates a new network topology and randomly spread nodes capable to give positive answers.

The query cost was estimated, using the overhead metrics for each successfully query. The data is obtained by counting, using the simulation environment.

The friends relationships are modeled by the graph G connectivity, where each node has its neighbour nodes. Each graph G = (V, E) is defined by a set of vertices V (which represent the network users) and a set of links between the vertices E (representing the friendship links between each user pair).
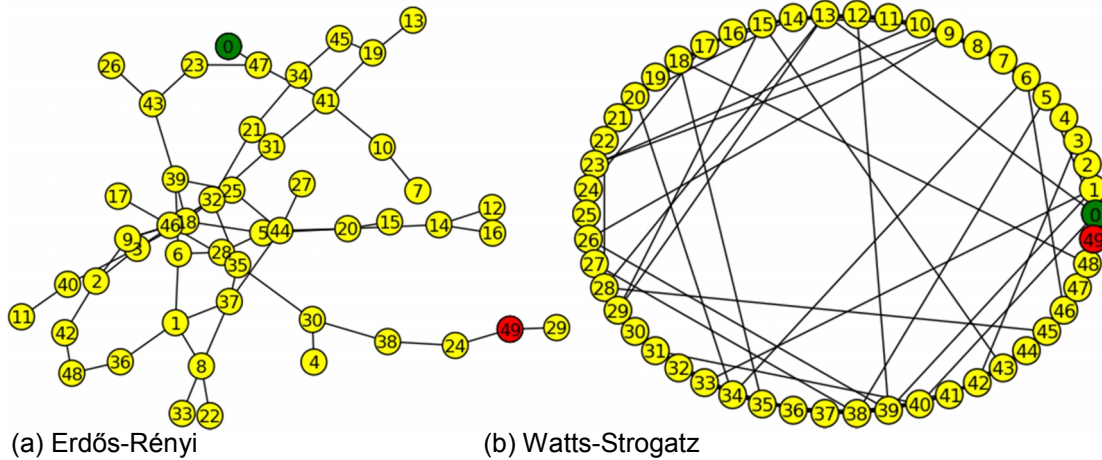

(a) Erdős-Rényi   (b) Watts-Strogatz

Fig. 2. Topologies graphs

The evaluation uses several topologies, depicted in Figs.1-2 corresponding to Barabasi-Albert, Erdős-Rényi and Watts-Strogatz graph models (Sala, et al., 2010). The depicted networks (Figs.1-2) have just 50 nodes, to improve graph readability.

The developed high-level simulation environment uses the package NetworkX[1], to store a new social network graph topology for each searching process. Users capable to successfully answer the query are random distributed over the network, with a density between 1% and 80%. In all the scenarios, it is assumed that at least one user knows the query answer. The overhead is acquired during the simulation and for each topology 2000 runs are iterated.
The first simulation scenario uses the Erdős-Rényi graph model (Fig. 2(a)) with a probability for edge creation of *p=0.01* (small p values causes sparse graphs). The second simulation scenario uses the Barabasi-Albert graph model (Fig. 1), where the graph is grown by attaching a single edge from a new node to existing ones. Finally, the last scenario uses the Watts-Strogatz small-world graph model (Fig. 2(b)), where each node is connected to nearest neighbours (*3% nodes*) in ring topology.

Results and Discussion. The results obtained by the simulation are depicted in Fig.3. The overhead data is hop aggregated, to expose the friendship level between the initiator user and the user that knows the answer. Low hop values represent *close friends* and as the hop raises the level of acquaintance decreases. The collected data is statistical analyzed and the resulting values rep- resent the average performance. The graph depicted in Fig. 3 compares flooding and ESC for different network topologies.

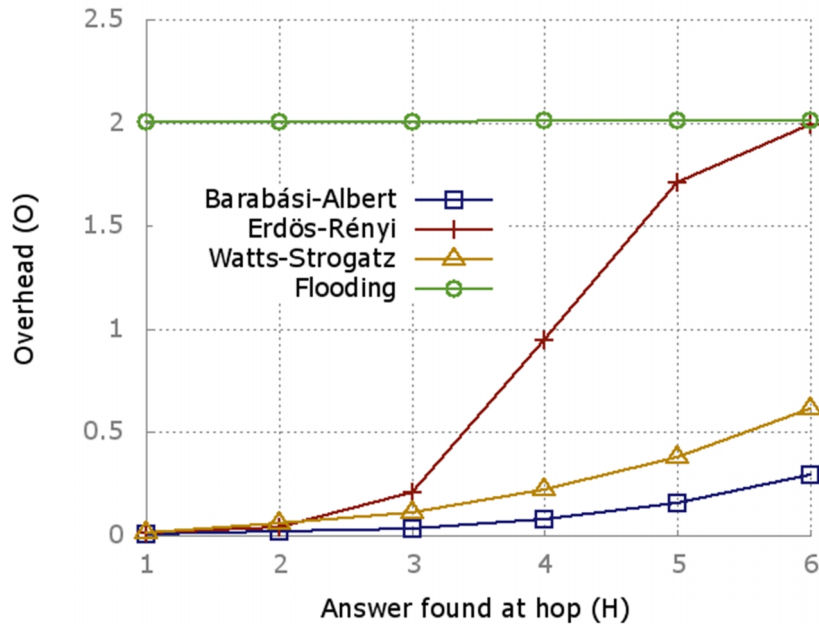[1] http://networkx.github.io/

Fig. 3. Topology influence in ESC

Flooding has high overheads even when the query can be answered by the closest friends, since all nodes will be contacted. The graph represents the flood performance, corresponding to forwarding searching messages to all nodes and sends also the answered messages. The flooding line will be positioned at overhead ≈ 2, corresponding to 2 × N messages.

In the case of ESC, the nodes will wait for a successfully answer (user intervention), before any node forward the query. Resulting in less query overhead (for any of the topologies), because users that know the answer will stop the query process. The delay introduced by network communications can be ignored when compared with the time taken by the users to act, and no additional latency considerations are done.

5. Conclusions

The outcomes of this research, provides a clear indication that expanded social circles (ESC) reduce the overhead, required to query users over social networks.

The results (Fig. 3) show that it is not efficient to bother everyone, when the neighbours (friends) can give a successful answer. The overhead savings are maximum when the closest friends circles can give successful answers, which is very frequent for a group of friends.

Results show that the proposed technique tolerates topology change, since all the explored topologies have similar overhead evolutions.

Future work. Additional research must be done to explore how to mitigate the spread of "*bad answers*" (misinformation), to minimize the number of wrong answers sent from specific friends circles. The query process can be extended to support access lists to avoid private information to be revealed. To get more realistic results, the ESC should be tested using datasets from online social networks.

References

Bjorklund, T. A., Gotz, M., & Gehrke, J. (2010). Search in social networks with access control. *Proceedings of the 2nd International Workshop on Keyword Search on Structured Data* (pp. 4:1--4:6). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/1868366.1868370

Crecelius, T., & Schenkel, R. (2012). Pay-as-you-go maintenance of precomputed nearest neighbors in large graphs. *Proceedings of the 21st ACM international conference on Information and knowledge management* (pp. 952-961). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2396761.2396881

Duchateau, F. (2011). Who can best answer a query in my social network? *Data Engineering Workshops (ICDEW), 2011 IEEE 27th International Conference on*, (pp. 218-223).

Lane, N., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., & Campbell, A. (2010). A survey of mobile phone sensing. *Communications Magazine, IEEE, 48*(9), 140-150.

Lima, R., Baquero, C., & Miranda, H. (2012). Stopping ongoing broadcasts in large MANETs. *Proceedings of the 1st European Workshop on AppRoaches to MObiquiTous Resilience* (pp. 4:1--4:5). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2222436.2222440

Lund, K., & Zhang, Y. (2011). The observation of tolerance in a social network model. *Proceedings of the 2011 Workshop on Agent-Directed Simulation* (pp. 57-63). San Diego, CA, USA: Society for Computer Simulation International. Retrieved from http://dl.acm.org/citation.cfm?id=2048355.2048363

Miluzzo, E., Lane, N. D., Eisenman, S. B., & Campbell, A. T. (2007). CenceMe: injecting sensing presence into social networking applications. *Proceedings of the 2nd European conference on Smart sensing and context* (pp. 1-28). Berlin, Heidelberg: Springer-Verlag. Retrieved from http://dl.acm.org/citation.cfm?id=1775377.1775379

Park, I., & Pu, I. (2007, march). Energy Efficient Expanding Ring Search. *Modelling Simulation, 2007. AMS '07. First Asia International Conference on*, (pp. 198-199).

Radlinski, F., & Dumais, S. (2006). Improving personalized web search using result diversification. *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 691-692). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/1148170.1148320

Reda, A., Park, Y., Tiwari, M., Posse, C., & Shah, S. (2012). Metaphor: a system for related search recommendations. *Proceedings of the 21st ACM international conference on Information and knowledge management* (pp. 664-673). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2396761.2396847

Sala, A., Cao, L., Wilson, C., Zablit, R., Zheng, H., & Zhao, B. Y. (2010). Measurement-calibrated graph models for social network experiments. *Proceedings of the 19th international conference on World wide web* (pp. 861-870). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/1772690.1772778

Sarwat, M., Elnikety, S., He, Y., & Kliot, G. (2012). Horton: Online Query Execution Engine for Large Distributed Graphs. *Proceedings of the 2012 IEEE 28th International Conference on Data Engineering* (pp. 1289-1292). Washington, DC, USA: IEEE Computer Society. Retrieved from http://dx.doi.org/10.1109/ICDE.2012.129

Xiang, R., Neville, J., & Rogati, M. (2010). Modeling relationship strength in online social networks. *Proceedings of the 19th international conference on World wide web* (pp. 981-990). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/1772690.1772790

Yang, X., Steck, H., & Liu, Y. (2012). Circle-based recommendation in online social networks. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1267-1275). New York, NY, USA: ACM. Retrieved from http://doi.acm.org/10.1145/2339530.2339728