

FriendRouter - Real-Time Path Finder in Social Networks

Wladston Viana, Mirella M. Moro
Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
{wladston,mirella}@dcc.ufmg.br

ABSTRACT

Online social networks have become a platform for running and optimizing classical algorithms. Here, we introduce a tool for finding paths between social network users in real-time, a task that classical solutions are not tailored for.

Categories and Subject Descriptors

H.3.5 [Online Information Services]: Web-based services—*Social networks, Path search algorithms*

Keywords

Social networks, Path search algorithms

1. INTRODUCTION

With the rise of online social networks, studies have tried to confirm the six degrees of separation theory and confirmed the average distance to 4.67 degrees in Twitter [1] and 3.74 degrees in Facebook [5]. Even with so small distances, there is currently no tool to find paths between pairs of users of those social networks in real-time. Classical solutions include the Dijkstra algorithm [2], and the heuristics Best-First-Search [4] and A* [3]. However, such algorithms perform an extensive search in the graph, which is impractical due to the large size of the social graph and the access restrictions imposed by the social networks' owners.

This work introduces FriendRouter, a real-time path finder in Social networks. It finds paths expanding typically less than 40 nodes of the graph, complying to the access restrictions and providing a performance gain of about 10^5 times if compared to Dijkstra's algorithm.

2. FRIENDROUTER OVERVIEW

FriendRouter uses a bidirectional search algorithm based on A* with an heuristic (optimal path not guaranteed) to find short paths (close to the average minimum separation) between users. Its biggest contribution is its ability to dramatically minimize the access on the graph, making real-time path finding over online social networks possible.

The algorithm starts with a backwards search, in which only nodes geographically closer (i.e. shorter distance between their reported home city) to the final node are considered. A minimum of 1400 nodes are obtained, generating

Table 1: Number of users explored during path finding

Path	Explored nodes		
	from target	from origin	Total
BillGates>Bill_Gross>GreylockVC>joe_hellerstein>seemohan	29	7	36
shakira>aplusk>Padmasree>nitya>AlonHalevy	6	2	8
BillGates>CERN>EPFLen>pirroh>ailamaki	2	14	16
LeoDiCaprio>aplusk>levie>joe_hellerstein>seemohan	31	7	38

the goalset (a set of nodes connected to the destination, all belonging to the same cluster) with people who live in the same location and its surroundings. Afterward, the forward search starts, using a cost function $f(x) = g(x) + h(x)$, where $g(x)$ is the number of hops from the source to node x , and $h(x)$ is a heuristic penalty based on the node's in-degree, out-degree and geographical distance to the target-node. At each step, the node with the minimum cost (given by $f(x)$) is explored, until a node from the goalset is found. The objective is to choose nodes with higher connectivity which are close to the destination node, thus directing the search to a local cluster where it can be completed with less effort.

3. RESULTS AND CONCLUSION

Our experiment uses real data from Twitter and shows that FriendRouter finds paths between generic pairs of users in the network. Dijkstra and BFS require reading a considerable part of the graph when the nodes are well separated, thus being impossible to run over Twitter. The results emphasize FriendRouter's advantage: the search is performed in an optimized fashion, expanding typically less than 40 nodes, which represents a performance gain of about 10^5 times if compared to the optimum Dijkstra algorithm. Table 1 illustrates test cases.

In conclusion, we have showed that geographical data can be very useful as a heuristic for path finding in social networks. The poster presentation will detail the algorithms employed and show more evaluation results including statistics on the heuristic elements. FriendRouter application is available online at <http://friendrouter.com>.

4. REFERENCES

- [1] A. Cheng. Six degrees of separation, twitter style. In <http://www.sysomos.com/insidetwitter/sixdegrees/>, 2010.
- [2] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, 1959.
- [3] P. Hart, N. Nilsson, and B. Raphael. A formal basis for the heuristic determination of minimum cost path. *IEEE Trans. Systems Science and Cybernetics*, 2(2):100–107, 1968.
- [4] J. Pearl. *Heuristics: intelligent search strategies for computer problem solving*. Addison-Wesley, 1984.
- [5] J. Ugander et al. The anatomy of the facebook social graph. *CoRR*, abs/1111.4503, 2011.