Poster: Spotting Suspicious Reviews via (Quasi-)clique Extraction

Paras Jain, Shang-Tse Chen, Mozhgan Azimpourkivi†, Duen Horng Chau, Bogdan Carbunar†
Georgia Tech, †Florida International University
{paras, schen351, polo}@gatech.edu, †{mazim003, carbunar}@cs.fiu.edu

***Abstract*—How to tell if a review is real or fake? What does the underworld of fraudulent reviewing look like? Detecting suspicious reviews has become a major issue for many online services. We propose the use of a clique-finding approach to discover well-organized suspicious reviewers. From a Yelp dataset with over one million reviews, we construct multiple *Reviewer Similarity* graphs to link users that have unusually similar behavior: two reviewers are connected in the graph if they have reviewed the same set of venues within a few days. From these graphs, our algorithms extracted many large cliques and quasi-cliques, the largest one containing a striking 11 users who coordinated their review activities in identical ways. Among the detected cliques, a large portion contain *Yelp Scouts* who are paid by Yelp to review venues in new areas. Our work sheds light on their little-known operation.**

I. Introduction

Review-centric online services like Yelp[1] and TripAdvisor crowdsource the job of reviewing businesses. The popularity and influence of reviews make such sites ideal targets for malicious behaviors: businesses commission fraudulent reviews to artificially boost their ratings. An estimated 16% of Yelp restaurant reviews are fraudulent [1].

Identifying suspicious review behaviors is critical to maintaining the integrity of online services and protecting their users. However, this task is challenging, as fraudsters' strategies can change rapidly. Crowdsourcing services such as Freelancer, Fiverr and Amazon Mechanical Turk are exploited for recruiting experienced review writers at a massive scale for nefarious purposes [2].

Recent research started to investigate network-based techniques for uncovering organized fraud by analyzing the link structures among potential fraudsters. For example, NetProbe uses an inference algorithm to find "near bipartite cores" formed among fraudsters and their accomplices on eBay [3]. More recently, Vlasselaer et al. find rectangles in bipartite graphs to detect social security fraud [4].

Interestingly, even though cliques[2] and quasi-cliques[3] have long been hinted as one of the strongest tell-tale signs of fraud, no prior work has studied if they indeed exist in online review websites like Yelp, where we can create a graph where each node represents a user, and an edge connects two users if they have reviewed common venues.

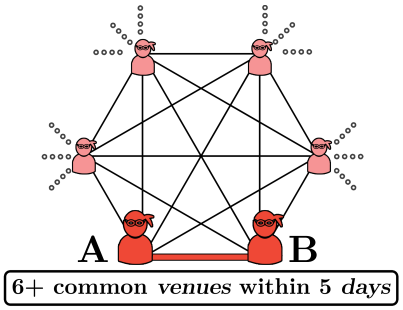

Fig. 1: Example suspicious clique formed among reviewers, found in a $(6, 5)$-graph extracted from the Yelp data. Two reviewers are linked when they have reviewed the same 6 venues (or more) within 5 days.

Our first research goal is to mine such graphs for cliques and quasi-cliques to verify the hypotheses from literature. Our secondary goal is to study if such cliques are indeed suspicious. Or relatedly, whether the (quasi-)clique structure is a strong indicator for fraud, whether there may be false convictions (e.g., any "good" cliques?), and if so, whether they are common and what approaches can reduce them.

We describe our preliminary results, which show that even from public data provided by Yelp, we can find large cliques that involve as many as 11 users—intuitively, this means that every possible pair of users (among these 11) reviewed multiple common venues within only a few days apart (see Figure 1). In practice, we should rarely see cliques of such large sizes. Possibly, the only legitimate setting for that to happen is that those 11 people are close friends or family members who always go to the same places together and also write reviews together!

II. Extracting (Quasi-)Cliques

The Yelp Review Dataset. We use the dataset from the *Yelp Dataset Challenge* [4], which contains 42,153 venues and 1,125,458 reviews. Yelp did not specify whether they filtered the dataset but it would be reasonable to assume that they have only included publicly listed reviews in the dataset.

Building K-D Graphs to Uncover Suspicious Links. We extract a set of (k, d)-graphs from the raw Yelp data, originally formatted as a list of JSON objects, varying the k and d

[1] http://www.yelp.com

[2] A complete sub-graph

[3] Synonymous with *pseudo cliques*

[4] http://www.yelp.com/dataset_challenge

k	Minimum number of commonly venues reviewed (by two users)
d	Maximum number of days between two reviews (written for the same venue)
Node	A Yelp Reviewer
Edge	Connects two users who reviewed $\geq$ k same venues within d days

TABLE I: (k, d)-graph definition (Reviewer Similarity Graph).

parameters. We define a (k, d)-graph to be an undirected graph, where vertices represent users, and an edge[5] exists between two vertices if the corresponding users reviewed at least k venues in common, and the reviews for each venue were posted at most d days apart (see Table I).

(Quasi-)clique extraction. We extract cliques and quasi-cliques from the set of (k, d)-graphs. Cliques are complete sub-graphs of undirected graphs. Quasi-cliques are sub-graphs with edge densities[7] no less than a fixed threshold[8] [5].

Identifying cliques is NP-hard. The Bron-Kerbosch algorithm finds maximal cliques and is based on the Branch-and-Bound technique. Most real-world datasets produce sparse graph, allowing Bron-Kerbosch to find maximal cliques faster than the theoretical worst case bound [6]. Suspiciously, large cliques of up to size 11 were found in the Yelp dataset (see Table II). Larger cliques with higher k (more venues) and lower d values (tighter time bound) are more suspicious.

To extract quasi-cliques, we utilize the method presented by Uno [5] which uses a greedy method to add nodes to the current quasi-clique, such that the edge density of the quasi-clique is greater than the threshold. Quasi-cliques of size 11 and 12 were found (see Table III).

***(k, d)*-graph**	3,5	3,6	3,8	4,5	4,6	5,5	5,6	6,5
9–clique	112	152	1040	29	73	13	28	10
10–clique	22	25	290	3	13	1	3	**1**
11–clique	**2**	**2**	**50**	—	**1**	—	—	—

TABLE II: Counts of large suspicious cliques, of sizes 9, 10, and 11, found in select (k, d)-graphs[6]. The most suspicious cliques are highlighted in red, due to large sizes, higher k (more venues) and lower d values (tighter time bound).

***(k, d)*-graph**	6,5	6,8	7,5	7,8	8,5	8,8	9,5
9–quasiclique	144	649	94	351	42	227	8
10–quasiclique	44	315	33	134	12	84	—
11–quasiclique	**7**	**100**	**4**	**33**	—	**15**	—
12–quasiclique	**1**	**20**	—	**4**	—	**1**	—

TABLE III: Counts of large suspicious quasi-cliques, of sizes 7, 8, 9, 10, 11 and 12 found in select (k, d)-graphs[6]. Suspicious quasi-cliques highlighted in red.

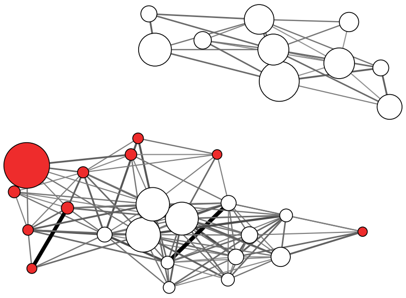

Fig. 2: Graph of combined cliques in a weighted $(6, 5)$-graph (weighted by size of the intersection of friends). Larger nodes represent more reviews. Red nodes are Yelp Scouts while white nodes are regular users. Scouts are tightly clustered and appear to associate with other Scouts.

We manually inspected some flagged users, and were surprised that they are *Yelp Scouts*[9] who are paid by Yelp to review venues in new areas. In the $(6, 5)$-graph, 31% of users were/are Yelp Scouts. This is a significant discovery; no prior study has revealed how these Scouts operate, how they choose which venues to visit, and what kinds of reviews they write (positive or negative)? Our work sheds light on these little-known activities, which are highly organized both in timing and in venue selection. While they may not be suspicious, they are certainly unnatural, and possibly controversial!

III. Conclusions & Next Steps

It is alarming to find large cliques from the Yelp data, which are likely suspicious. Still, some might not be. And it is critical that we devise methods that reduce the false alarm rates to the minimum possible, as it is greatly harmful to wrongly convict a good user. We plan to incorporate other rich signals from the Yelp data to help with this, such as by analyzing review text, and the spatial and temporal relationships among reviewed venues (e.g., it would impossible for a user to visit a venue in the US and another in Asia on the same day).

References

[1] M. Luca and G. Zervas, "Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud," *Harvard Business School NOM*, pp. 1–25, 2014. [Online]. Available: http://businessinnovation.berkeley.edu/WilliamsonSeminar/luca092613.pdf

[2] G. Lubin, "The Illegal Way To Improve Your Rating On Yelp. Business Insider," 2012. [Online]. Available: http://www.businessinsider.com/the-illegal-way-to-improve-your-rating-on-yelp-2012-2

[3] S. Pandit, D. H. Chau, S. Wang, and C. Faloutsos, "Netprobe: a fast and scalable system for fraud detection in online auction networks," in *WWW*. ACM, 2007, pp. 201–210.

[4] V. Van Vlasselaer, L. Akoglu, T. Eliassi-Rad, M. Snoeck, and B. Baesens, "Guilt-by-constellation: Fraud detection by suspicious clique memberships," in *HICSS*, 2015.

[5] T. Uno, "An efficient algorithm for solving pseudo clique enumeration problem," *Algorithmica*, vol. 56, no. 1, pp. 3–16, 2010.

[6] C. Bron and J. Kerbosch, "Algorithm 457: finding all cliques of an undirected graph," *Commun ACM*, vol. 16, no. 9, pp. 575–577, 1973.

[5] Edges can be weighted with a calculated similarity score between users

[6] Abbreviated table—some (k, d)-graphs not displayed.

[7] Number of edges that exist in sub-graph over number of edges in a complete graph with same number of vertices.

[8] $\theta = 0.90$

[9] A user's Yelp Scout status is determined from a badge on their profile