Cascading Behavior in Yelp Reviews

Muhammad Raza Khan
University of California, Berkeley
mraza@berkeley.edu

ABSTRACT
Social media has changed the landscape of marketing and consumer research as the adoption and promotion of businesses is becoming more and more dependent on how the customers are interacting and feeling about the business on platforms like Facebook, Twitter, Yelp etc. Social review websites like Yelp have become an important source of information about different businesses. Social influence on these online platforms can result in individuals adopting or promoting ideas and actions resulting in information cascades. Research on information cascades have been gaining popularity over the last few years but most of the research has been focused on platforms like Twitter and Facebook. Research on the adoption or promotion of product using cascades can help determine important latent patterns of social influence.

In this work, we have analyzed the spread of information i.e. cascades in Yelp across different cities in Europe and North America. We have enumerated and analyzed different cascades topologies that occur in the Yelp social networks. Some of our significant findings include the presence of a significant number of cascades in Yelp reviews indicating the importance of social influence, heavy tailed distribution of cascades and possibility to accurately predict the size of cascades on the basis of initial reviews. In addition, we have also found that the characteristics of the non root nodes and the non root reviews are much more important type of feature as compared to the properties of the root nodes and root reviews of the cascade.

These findings can help social scientists to analyze customer behavior across different cities in a much more systematic way. Furthermore, it can also help the businesses in a city to figure out different consumer trends and hence improve their processes and offerings.

1 INTRODUCTION
Social media websites like Facebook and Twitter etc. have become an important source of information to analyze human behavior in different circumstances. Before the popularity of these social media websites, the research on human behavior in many cases was constrained by the lack of availability of data. However, these social media websites have become the biggest observatory of human behavior, as a result of which quite a significant number of research projects on human behavior analysis have employed social media data. People express their personal feelings and concerns and at the same time review businesses and products on these social media websites. On the other end, businesses and companies have been using these social media outlets to promote their products, customer service and quality. The way customers deal with these promotions or express their views about these companies can tell a lot of information about these businesses. The research on social media analytics is primarily dominated by the work correlating business performances with the reviews, however the impact of the reviews of the individuals on others has been a relatively less popular theme. The impact of the reviews of an individual activity in the social network over other individuals' activity is the broader theme of this work.

Social networks impact human behavior at different levels in different contexts and it has been difficult to estimate the social influence on individual's behavior specially when it comes to buying something or dealing with a business. To quote Subramani and Rajagopalan [14], "there needs to be a greater understanding of the contexts in which viral marketing strategy works and the characteristics of products and services for which it is most effective. ... What is missing is an analysis of viral marketing that highlights systematic patterns in the nature of knowledge-sharing and persuasion by influencers and responses by recipients in online social networks". Part of the problem in effectively analyzing the influence of social networks has been the lack of availability of data, however the growing research on Facebook and Twitter data indicates that this is becoming less of a problem. The other half of the problem is that the impact of users on other users can be a complex function of different latent features. The influence of influencers on others can be different even with in a social network.
depending on the geographic location or other characteristics of the users. Information cascades are a common phenomena in the social networks in which the user adopt a new idea or perform an action under the influence or recommendation of other users. Information cascades are an ideal tool to analyze the impact of social influence. Furthermore, they can also help in the analysis of important factors behind social influence in different contexts. In this work, we have analyzed millions of reviews over the social review website Yelp from million of users across different cities in Europe and North America.

1.1 Present Work
Using the data officially provided by Yelp¹, we explore the following questions in this research paper.

- What kind of information cascades frequently arise in Yelp reviews across different cities?
- What are the significant features of cascades across different cities?
- Given information about a few early reviews, can one predict the longevity of the influence? (i.e. Whether these reviews are a part of a big or small cascade?)

Using Yelp reviews and Yelp network data from different cities we have tried to have a systematic understanding of patterns of influence across different cities exploring what sort of generic patterns can be deduced.

1.2 Summary of results
There are four substantive and one methodological contributions of this study. Substantively, we (1) develop a richer understanding of what drives the diffusion of information in Yelp reviews; (2) construct a supervised learner that can predict, to varying degrees of accuracy depending on the city context, the likelihood than a cascade is going to be a big one or short one; (3) analyze the structural properties of the cascades across different cities and indirectly analyze the properties of the underlying social network; and (4) analyze the features that can help in predictability of the virality of the cascade. To our knowledge, the analysis of information diffusion at this scale across different cultural contexts has not been performed earlier.

Methodologically, we have developed a framework to (1) enumerate and analyze cascades in a scalable and efficient manner; (2) extract the features of cascades (3) and predict whether the cascade is a long one or short one.

The analysis of cascades can be beneficial for both the business owners and the review websites (Yelp in this case) as it can unearth latent trends of consumer behaviors in different cultures which can in turn help the business owners to improve their processes.

2 RELATED WORK
Our work builds on several distinct strands in the academic literature. The first is concerned with the general diffusion of information in social networks. This has been a popular topic of research over the last decade. [2] et al. have tried to analyze the cascading behavior in Flickr and their key finding has been that growth of cascades in Flickr can be much faster as compared to some epidemics like infectious diseases indicating that social networks can be a very effective medium for transmission of information. [4] et al. have tried to compare and contrast the growth of two cascades on Facebook. Consistent with other studies they have found out that small proportion of content generates cascades of non-trivial size and depth. However, their finding that different cascades can reach the same level of popularity through different network topologies and different source and chain characteristics is of special relevance for us. This is because, we also wanted to find out what are the frequent cascades shapes in different contexts and what part do the features of the source node and the features of the chains play in the growth of the cascade.

[10] have tried to analyze the impact of blogs on each other. The most relevant literature to our work is the work on patterns of influence in recommendation networks [11], [9]. In these works, the authors have tried to analyze the patterns of social influence in recommendation networks. Their work and our work is quite similar in methods but we want to find the patterns of influence specifically in the Yelp social network. Furthermore, we want to identify important determinants for the growth of cascades in social networks. Our work also confirms the finding of the [3] as we also

| City            | No of businesses | No of reviews | No of users | Total no of cascades | CL90: Cascades Length | Long Cascades |
|-----------------|------------------|---------------|-------------|----------------------|-----------------------|---------------|
| Karlsruhe       | 2,965            | 33,888        | 12,266      | 8                    | 2.2                   | 1             |
| Edinburgh       | 3,539            | 50,781        | 12,261      | 300                  | 6.0                   | 28            |
| Montreal        | 6,668            | 131,927       | 52,160      | 561                  | 5.0                   | 46            |
| Pittsburgh      | 8,091            | 212,107       | 76,577      | 823                  | 8.0                   | 63            |
| Charlotte       | 10,177           | 282,884       | 99,345      | 896                  | 8.0                   | 69            |
| Madison         | 3,899            | 102,140       | 38,698      | 449                  | 7.0                   | 34            |
| Urbana Champaign| 1,556            | 33,176        | 15,431      | 185                  | 4.1                   | 15            |
| Cleveland       | 9,966            | 228,990       | 81,176      | 108                  | 2.0                   | 6             |
| Waterloo        | 24,507           | 585,119       | 126,970     | 462                  | 2.0                   | 24            |
| Las Vegas       | 28,214           | 1,846,944     | 627,609     | 5,665                | 4.0                   | 312           |
| Phoenix         | 43,492           | 1,556,500     | 429,634     | 3,267                | 3.0                   | 188           |

¹https://www.yelp.com/dataset_challenge/

Notes: CL90 indicates the 90th percentile for the length of all the cascades in a particular city
Last column indicates the number of long cascades in each of the city
find out that cascades can be predicted quite accurately using the information about the seed nodes and initial nodes of the cascades.

Another set of literature related to this research includes the work that explores the impact of different structural characteristics of social networks to model product adoption and diffusion of information. Bakshy et al. found that the adoption rate in the Second Life social network increases as the number of adopting friends increases and the effect of the number of adopting friends depends on the connectivity of the individual user [1]. In another related strand of literature, the researchers have tried to analyze the impact of social networks on product adoption across different social networks (For instance, Khan et al. have tried to explore the impact of different features like the alter of an ego node for the purpose of product adoption in developing countries [8]) and there has been mixed evidence about the impact of social network. Though the consensus is that the social networks generally do play a role in product adoption, the extent of impact of these social networks varies. Similarly, Ugander et al. found that probability of contagion is more tightly connected to the structural components in an individual’s network than the actual size of the network [15] while Romero et al. have tried to analyze the interplay between social and topical structure for predicting the growth of hashtags [13]. Their work is related to our research as we also try to analyze the impact of different features on the growth of the cascades. There has been limited work on diffusion of information in Yelp reviews. For instance Gee et al. [6] have tried to explore whether a modified herdimg model better explains the existence of cascades in yelp or not. Though the dataset that they have being using is quite similar to what we are doing but we are more focused on the features that can help predicting the growth of the cascades across different cultures.

One distinguishing feature of this study is that we have analyzed the spread of cascades on Yelp which can be slightly different from the platforms like Facebook and Twitter where the users may not have to purchase anything. To conclude, this work is different from other works as it deals with the Yelp social network and is focused on finding systematic patterns of influence across different cities. Relative to the studies mentioned in this section, our study confirms the existence and heavy tailed nature of cascades. But by focusing on the different type of features that can help in distinguishing between long and short cascades we have tried to narrow down on the factors that determine the popularity or virality of cascades generically across different cities.

3 DATA AND CONTEXT
For this particular study, we have been using the data provided by the Yelp under its Dataset challenge initiative⁷. As our basic aim in this research was to see the underlying patterns of social influence across different cities we decided to use the data both from the Round 8 and Round 9 of the challenge. The data in total consists of around 6 million reviews and 1.5 million tips for more than a million users and 150 thousand users across 11 different cities in the Europe and North America. In addition, to the reviews and tips the dataset also contains the business attributes like hours, parking availability and ambiance etc. as well. Though the dataset is quite big in total but the number of reviews is quite low in the earlier years as shown in the Figure 1. For the purpose of this research, we have focused on cascades resulting from reviews and tips for the year 2015, though a similar analysis can be easily done for any other year provided sufficient data is present. Some of the summary statistics for each of the cities in our analysis can be found in the Table 1.

To find out cascades in the data we had to find out the reviews and tips that result in future review and tips. Though the behavior of the person deciding to deal with a business can be influenced by many different factors like visibility of the reviews, explicit referral from an acquaintance; we are adopting a more strict approach such that only the friends of a user can be influenced by the reviews or tips of the users. In other words, an edge \((u, v)\) will be part of a cascade for a business \(b\) if the user \(u\) writes a review or tip at time \(t\) and the user \(v\) write a review or tip at the time \(t'\) such that \(t' > t\) and \(v \in N(u)\), where \(N(u)\) specifies the neighboring nodes of \(u\). All the edges that specify this property will constitute a cascade for the business \(b\).

To frame the problem, as a prediction task, we calculate the length of each cascade for a business as the number of nodes or users who participate in the cascade either by writing a review or tip for the business. Furthermore, we categorized the cascades as long or short cascades as explained below

- **Long Cascades**: If the length of a cascade is greater than or equal to the 90th percentile of the length of all the cascades of the city. The 90th percentile of the lengths of all the cascades for a city is shown in the second last column of the table 1.
- **Short Cascades**: If the length of a cascade is less than the 90th percentile of the length of all the cascades of the city.

As our primary interest in this work was to compare long vs short cascades in a city, we had to eliminate Karlsruhe, Waterloo and Cleveland from our analysis as most of the cascades in these cities were short cascades. However, given enough data our models and analysis can be easily extended to these cities as well.

⁷https://www.yelp.com/dataset_challenge/
4 STRUCTURAL ANALYSIS OF CASCADES

Size Distribution of Cascades
As the first step in our analysis, we analyze the distribution of the size of cascades in different cities. The distribution of size of cascades for different cities is shown in Fig. 2. It is obvious from the Fig. 2 that cascades in all of the cities follow a heavy tailed distribution. Most of the cascades are small but large cascades do happen. Furthermore, this figure also shows that the cascade distribution across different cities has a power-law exponent in the range $-2.15$ and $-1.68$. Las Vegas and Pheonix have the highest number of large cascades but large cascades do exist for other cities as well.

Frequent Cascade Topologies
As the next step in the structural analysis of cascades we wanted to see that what kind of cascades frequently arise in Yelp reviews across different cities. As discussed in [11], no polynomial time graph isomorphism algorithms exist and for the purpose of enumerating graphs one has to generate an efficiently computable signature of a graph and count the number of graphs with similar signatures. [11] have proposed a multiple level signature for frequent subgraph mining. The signature used for hashing the graphs consists of a) Number of Nodes in the graph; b) Number of edges in the graph; c) Sorted in degree sequence of each of the node; and d) Sorted out degree sequence of each node. Using this technique we count the frequency of frequent cascades for each of the cities. Rank and frequency of the frequent cascade topologies for each of the cities is shown in the Table 2. As expected, the most common cascade i.e. G1 as shown in the Table 2 corresponds to the case in which a person participates in a cascade under the influence of only one other node. This corresponds to more than fifty percent of the cascades across all the cities and in some cases this particular type of cascade represents more than eighty percent of all the cascades of a city (e.g. Karlsruhe). Comparing G2 and G10, we can see that receiving the recommendation or influence by more than once channel increases the likelihood of participation but at the same time, G3 and G11 shows that the presence of “influential” nodes also makes cascades longer. The cascades like G5 occur when both the user and the friend f either review or write tip for the same business id on the same day.

Longest Cascades
On a similar line of exploration, we also visualize the longest cascade in each of the cities. The result of this exploration are shown in the Figure 3. The presence of some influential central nodes is obvious in the cities like Las Vegas, Phoenix and Charlotte, the cities having higher number of business and reviewers. One clear exception to this trend is Waterloo where the size of the longest cascade is lower (90th percentile of the cascade length for each of the cities is shown in the table 1) as compared to the other cities but the number of business and reviewers on Yelp from waterloo is quite high.

5 PREDICTING CASCADE GROWTH

One of our main aim in this research was to see whether we can predict the growth of the cascade using a few initial reviews and tips. And as a second step, we wanted to analyze the features having high predictive power when it comes to predicting growth of cascades.

5.1 Experimental Design
Before going into the details of the feature engineering and cascades modeling, we want to specify the setup of our experiments in detail. First of all, we are trying to analyze the influence of a review or a tip made by the user on the activities of the members of the social network of that person. Though all of the yelp reviews are public and one can be influenced by the reviews and tips of the persons not in the friends list of that person, for the purpose of this research we consider only the friends of a person as a part of his social network. Yelp app and website has a specific interface of adding friends through Yelp, email or Facebook friends list 3. Secondly, the temporal ordering of reviews and tips by the users for

3https://www.yelp-support.com/article/How-do-I-add-friends-on-Yelp?l=en_US
a particular business gives us the edges of the cascades. The formulation of cascades as a prediction problem needs special attention because most of the cascades are small. [3] have summarized some of the issues and solution while modeling the growth of cascades. For the purpose of this study, we divide the cascades in each of the cities in two categories namely short cascades and big cascades where the cascades are categorized as big cascades if their length if greater than ninety percentile of the lengths of all the cascades of that city. The decision of ninetieth percentile may seem adhoc but it made sure that we are able to emphasize on the truly long cascades in a city. However, one impact of this decision was that we were having very few number of long cascades for Karlsruhe, Cleveland and Waterloo, hence we did not use the cascades from these cities in our predictive experiments.

The classification of cascades on the basis of percentiles may seem arbitrary but it makes sure that we are focusing on truly long cascades in each of the cities as the cascades in all of the cities are heavy tailed 2. A better experimental design would have been to model the growth of the cascades over time i.e. predicting whether the cascade would reach 2 * k users, given that it has reached k users (just like [3]) but our approach still sufficiently handles our objectives i.e. the study of predictability of cascades status (long vs short) and the features determining the cascades length.
Furthermore, equal number of long and short cascades were selected for each of the city.

5.2 Features Engineering
The table 3 shows the list of features used for predicting longevity of the cascades. Primarily the features are divided into five main categories namely: Business related features, Root node features, Non-root node features, Root review features, Non-root review features.

5.3 Classification and Model Selection
Using the features specified in the last section, we used a variety of supervised learning algorithms to classify the cascades as long or short cascades. Since our data has a large number of features relative to observations, we focus on learners that are robust to over-fitting, such as regularized logistic regression [16], gradient boosting [5]. Performance was comparable across these classifiers, although as expected these methods generally performed better than unregularized alternatives. To streamline the analysis summarized in the section 6, we report only the results from gradient boosting, which outperformed the other classifiers by a small margin. In all cases, we report the average accuracy across testing sets from 5-fold cross validation.

The level of a feature used as a decision node in the decision tree can be used to assess the relative importance of the feature to predict the longevity of the cascades. We use the feature importance calculated in this way to identify important feature categories as described in the section 6.

6 RESULTS AND DISCUSSION
As stated earlier, our first goal in this research was to evaluate whether cascades growth can be predicted on Yelp or not. To this end, he predictive accuracy and ROC curve (based on our best performing gradient boosting algorithm as described in the section 5) for predicting status (long vs short) of cascades across different cities is shown in the Figures 4a and 4b respectively.

In addition to calculating the predictive accuracy and ROC for the cascades in each of the cities, we also determine the importance of different type of features for each of the city as well. We have generated five different type of feature categories explained as follows.

- **Root Features**: All the features of the original node(i.e. person) who started the cascade are categorized as root node node features. Some example of the root node review features include the gender of the root node and the number of reviews the root node has contributed.

- **Non Root Features**: This category includes the features of all the non root nodes that contribute to the cascade. Some example of the features in this category include average number of friends for the non root nodes for a cascade, average number of reviews by all the non root nodes.

- **Root Review Features**: All the characteristics of the text of the root review (the review that started the cascade) are categorized as the root review features. For example, the number of “cool” votes received by the root review.

- **Non Root Review Features**: All the characteristics of the non root reviews contributing to the cascade are categorized under the non root review features. Total number of cool votes received by all the non root reviews is one such example.

- **Business Features**: The last category of features includes all the features for the business receiving the review or tips. Total number of reviews received by the node is one example of business related features.
More details of different features used in the analysis can be seen in the Table 3. The sentiment prediction (positivity) for the actual text of the reviews is performed using a model pre-trained on the dataset.
on the Yelp and Amazon reviews with ratings greater than being considered as positive reviews. Importance for different type of features is shown as violin plots [7] in the Figure 5. To understand which features are more important for prediction of cascades, we calculate the importance of each feature using gradient boosting classifier. The importance of features is the average of importance in each of the individual trees in the gradient boosting. Each of the trees calculates the importance of features using the information gain with intent to favor pure subtrees at each level of the tree.

Some interesting patterns can be observed in the results shown in the Figures 4, 5. For instance, cascades of the cities like Phoenix, Montreal and Edinburgh cannot be predicted very accurately while those of Pittsburgh and Madison can be predicted with quite high accuracy.

In all the cities, where the cascades can be predicted with higher accuracy, the features belonging to the non root nodes have higher predictive importance. Amongst the cities with low predictive results, only Edinburgh has got higher feature importance for the non root node category. In none of the cities where the cascades can be predicted with higher accuracy, the root nodes and root reviews have higher predictive importance which shows that the first reviewer and the first review may not be that important in the case of cascades in Yelp reviews or in other words, non influential nodes may start the cascades.

Top two features in each of the cities are shown in the Table 4. The variance in the yelping age of the non root reviewers is the top feature in most of the cities which indicates that the users who are active and relatively long term users of Yelp are more likely to be part of the cascades. Different type of votes that a review, reviewer or a business may receive is the dominant type of second most important feature.

### Table 4: Top predictive features

| City             | Most Important Feature                                                                 | Second Most Important Feature                                      | Predictive Accuracy (%) |
|------------------|----------------------------------------------------------------------------------------|---------------------------------------------------------------------|-------------------------|
| Edinburgh        | Variance in the yelping age of the non root reviewers                                  | No of useful votes for the root reviews                            | 57.89                   |
| Madison          | Variance in the yelping age of the non root reviewers                                  | No of hot compliments for the business                             | 89.7                    |
| Charlotte        | Variance in the yelping age of the non root reviewers                                  | Average star rating for the non root reviewers                      | 85.82                   |
| Pittsburgh       | Variance in the yelping age of the non root reviewers                                  | No of photo compliments for the business                           | 97.14                   |
| Urbana Champaign | Average number of elite ratings for the non root reviewers                             | No of useful votes on the reviews of the business                   | 75.33                   |
| Las Vegas        | Variance in the yelping age of the non root reviewers                                  | No of useful votes on the reviews of the business                   | 71.16                   |
| Phoenix          | Number of cool votes for the business                                                  | Average star rating for the non-root reviews                        | 62.0                    |
| Montreal         | Number of stars for the root node                                                      | No of cool compliments for the root node                            | 64.83                   |

### Figure 5: Importance of features across cities

In this paper, we have tried to analyze the cascades on Yelp in terms of the predictability of them being long or short. At the same time, we have tried to identify the important properties of these cascades and the properties of the nodes and the business involved in these cascades. Though, the predictive performance of our models vary across different cities but some notable conclusions can be drawn from our analysis. First of all, in most of the cities with the higher number of business and reviews the longevity of the cascades can be predicted quite accurately. Secondly, the characteristics of the non-root reviewers are a more powerful indicator of the popularity of the cascade as compared to other categories like root reviewer and root review characteristics. Lastly, the variance in the age of the non root node reviewers is the most important feature in most of the cities.

There have been very few attempts to systematically analyze the information flow or propagation on Yelp but these information cascades have the potential to help business owners and Yelp both to improve their services. Most of the focus of this research has been...
to identify the structural and nonstructural features of the cascades on Yelp and see the relationship between these features and length of the cascades. In future, we want to incorporate additional data sources like pictures and images of the businesses in our analysis. We also want to experiment with the recent deep learning based methods (e.g. [12]) to model the growth of cascades.

Right now, we are not imposing any time bound on the edges of the cascades which may imply that the successive edges in a cascade may be months apart. In our future work, we want to apply additional constraint of time bound on the successive edges in a cascade. Though our focus in this work has been on the cascades in Yelp social networks but a similar analysis can be extended to other types of social network provided significant data is available.

REFERENCES

[1] Eytan Bakshy, Brian Karrer, and Lada A. Adamic. 2009. Social Influence and the Diffusion of User-created Content. In Proceedings of the 10th ACM Conference on Electronic Commerce (EC ’09). ACM, New York, NY, USA, 325–334. https://doi.org/10.1145/1566374.1566421
[2] Meeyoung Cha, Alan Mislove, Ben Adams, and Krishna P Gummadi. 2008. Characterizing social cascades in flickr. In Proceedings of the first workshop on Online social networks. ACM, 13–18.
[3] Justin Cheng, Lada Adamic, P Alex Dow, Jon Michael Kleinberg, and Jure Leskovec. 2014. Can cascades be predicted?. In Proceedings of the 23rd international conference on World wide web. ACM, 925–936.
[4] P Alex Dow, Lada A Adamic, and Adrien Friggeri. 2013. The Anatomy of Large Facebook Cascades. In Seventh International AAAI Conference on Weblogs and Social Media.
[5] Jerome H Friedman. 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics (2001), 1189–1232.
[6] Grace Gee, Chris Lengerich, and E ONeill. 2011. Identifying Cascades in Yelp Reviews. Technical Report. Tech. rep., Stanford University.
[7] Jerry L Hintze and Ray D Nelson. 1998. Violin plots: a box plot-density trace synergism. The American Statistician 52, 2 (1998), 181–184.
[8] Muhammad R. Khan and Joshua E. Blumenstock. 2016. Predictors Without Borders: Behavioral Modeling of Product Adoption in Three Developing Countries. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ’16). ACM, New York, NY, USA, 145–154. https://doi.org/10.1145/2939672.2939710
[9] Jure Leskovec, Lada A Adamic, and Bernardo A Huberman. 2007. The dynamics of viral marketing. ACM Transactions on the Web (TWEB) 1, 1 (2007), 5.
[10] Jure Leskovec, Mary McGlohon, Christos Faloutsos, Natalie Glance, and Matthew Hurst. 2007. Patterns of cascading behavior in large blog graphs. In Proceedings of the 2007 SIAM international conference on data mining. SIAM, 551–556.
[11] Jure Leskovec, Ajit Singh, and Jon Kleinberg. 2006. Patterns of influence in a recommendation network. In Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, 380–389.
[12] Cheng Li, Jiaqi Ma, Xiaoxiao Guo, and Qiaozhu Mei. 2017. DeepCas: an End-to-end Predictor of Information Cascades. In Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 577–586.
[13] Daniel M Romero, Chenhao Tan, and Johan Ugander. 2013. On the Interplay between Social and Topical Structure. In Seventh International AAAI Conference on Weblogs and Social Media.
[14] Mani R. Subramani and Balaji Rajagopalan. 2003. Knowledge-sharing and Influence in Online Social Networks via Viral Marketing. Commun. ACM 46, 12 (Dec. 2003), 300–307. https://doi.org/10.1145/953460.953514
[15] Johan Ugander, Lars Backstrom, Cameron Marlow, and Jon Kleinberg. 2012. Structural diversity in social contagion. Proceedings of the National Academy of Sciences 109, 16 (2012), 5962–5966.
[16] Hui Zou and Trevor Hastie. 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, 2 (2005), 301–320.