Predicting Network Evolution through Temporal Twitter Snapshots for Paris Attacks of 2015

Max Berest[1], Ralucca Gera[2], Zachary Lukens[2] , Nicolas L. Martinez[3] , Ben McCaleb[3]

[1] Department of Computer Sciences, Naval Postgraduate School, Monterey, CA
[2] Department of Applied Mathematics, Naval Postgraduate School, Monterey, CA
[3] Department of Operations Research, Naval Postgraduate School, Monterey, CA

Abstract. As technology advances, modern networks rapidly evolve. Capturing the dynamic nature of networks and predicting their evolution has been a common focus in network science. This research investigates a social network's temporal evolution, and how metrics and descriptors during its creation compare a snap shot in time during the network's growth to the known state of the final network. As social media is a primary way of communication, Twitter data collection provide real traces for this study that focuses on the ability to determine if knowing network's early metrics provide an accurate prediction of the this final network. This can then be extended to monitor other similar events as they are happening. However, this does not generalize arbitrary social evens. Specifically, this research utilizes data from Twitter feeds regarding the Paris terrorist attacks ($\#ParisAttacks$) in November 2015, and focuses on the analysis of k-Core, Betweenness centrality, and community comparison as the network grows. The topology of the overall network after 24 hours from the time of the first post provides the known "end-state" that we compare against.

Keywords: network topology, Betweenness centrality, k-Core, distance in a network, communities, clustering, Twitter, $\#ParisAttacks$, network temporal growth.

1 Introduction

On the evening of 13 November 2015, six coordinated terrorist attacks were launched against the civilian populace in Paris, France 1.

These attacks killed 130 Parisians and wounded another 368. The city of lights lived in fear that night of continued violence, desperate to know if loved ones were still alive. The city was shut down with the citizens and tourists sheltering in homes and hotels. During this time, many of these individuals took to social media forums like Facebook and Twitter to communicate, find loved ones, and share news of the attacks that they were experiencing first hand. Through this messaging, the world became involved in a singular world event.

For over 24 hours, the social media platform Twitter was used by millions across the world to distribute news of the attacks. Many of these tweets shared

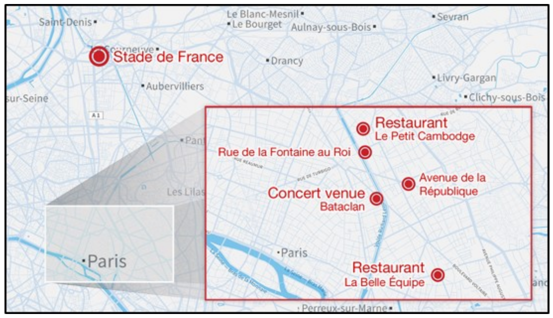


Fig. 1. The locations of the 6 attacks [20]

information, broadcasted a reaction, or responded to the attacks. Twitter account holders used the trending message qualifier *#ParisAttacks* to join their individual message to the overall collective response and discussion of the attacks.

These messages, or "tweets", are constructed of three parts in building an event based social network: author/sender, receiver, and time. An event's occurrence results in only a small group of Twitter users tweeting about the event until an informal trending topic handle is assigned that is soon adopted by everyone. Twitter data for the night of 13 November and the ensuing 24 hours recorded tweets referencing the trending topic *#ParisAttacks*. This handle provided the world of Twitter account users the ability to receive and pass on information and news regarding the event. After 24 hours there were more than 86,000 people receiving and then transmitting *#ParisAttacks*. Netlytic [8], a social network analysis tool, was used to collect the *#ParisAttacks* tweets. The resultant network from our data is shown in Figure 2.

This network of Twitter users communicating about a single event over time has the ability to provide insights on the evolution of information networks. Specifically, this paper studies the characteristics of the entire network at the 24^{th} hour of growth and attempts to identify at what period in time, beginning at hour 1, does the end-state network take its final form. The hypothesis of this paper speculates that when this network is observed over time that the development of its k-core will reveal the gradual framing of the main elements and characteristics of the end-state network. The observation of these characteristics can aid in further academic endeavors studying how networks grow over time.

The remaining of the paper is organized as follows. In Section 2 we introduce the relevant definitions and literature review for our study. In Section 3 we introduce our methodology, whose results and analysis on the three data sets are presented in Section 4. We then present the conclusion of our paper in Section 5.

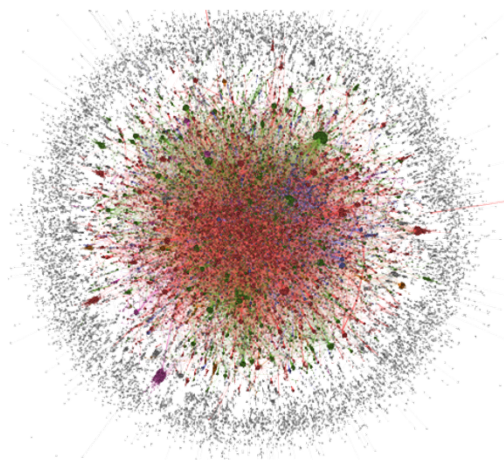


Fig. 2. End-state Network

2 Relevant Definitions and Literature review

In Analyzing network's structure, researchers have considered three different levels of analysis. The global level is studied through the macroscale type analysis (like degree distribution), the detailed level is through the microscale analysis (such as degree), and the mesoscale such as k-core and communities that we will use in the current research.

A key concept used in the current research is the k-core of a network. Graph theory introduces locally dense structures (such as components and cliques [12, 24, 9]) which have been further extended in complex networks to k-clans, k-clubs, k-clique-communities, diplex k-plexes or k-core since early 1900s [4, 14, 12, 16, 17, 18, 21, 25]. The k-core of a network is the subset of nodes that have k or more relationships in the k-core [23, 6]. For a review article on core-periphery structure see [5].

Some of the above mentioned locally dense structures were then used for defining communities in networks. While there is not a widely accepted definition of communities, traditional community detection researchers use Radicchi [19] definition of community as a general concept as a sub-network or subset of vertices with more internal edges to the community compared to edges between communities. This definition we will be using in the current research through the use of Louvain method [2] of maximizing modularity as it is one of the more popular community detection algorithms in network science. While the Louvain method is not deterministic, we used Gephi's implementation and varied the resolution until the maximum modularity value was achieved. Once these parameters were identified, the communities were defined. The modularity of a network's partition into communities measures the goodness of the particular partition by comparing it to a partition of a similar size (node and edge count) random network [15]. That is, the modularity is the result of summing $a_{ij} - \frac{k_i k_j}{2m}$ for all pairs (i, j) in the same community. Community detection is an extremely active research area with many modularity based algorithms.

The study of social media, including Twitter, and its associated network structures is a broad and developing area of research. The pervasiveness and generous amount of data in this area makes for most interesting academic research and consideration. As expected, researchers have considered Twitter analysis of disaster events, none that we could found on the $\#ParisAttacks$.

The devastating earthquake and tsunami that hit Japan in 2011 claimed the lives of thousands of people. The authors of [1] analyze tweets sent immediately after the disaster. The focus of their analysis is on the types of messages people sent based on their location in Japan.

The authors of [13] analyze Twitter activity related to the 2010 earthquake in Chile. As the earthquake began, social media platforms were used in a similar manner: to connect, share information, and form a collective conversation. The paper compares the way news versus rumors are spread about the earthquake via Twitter, and they show that it is possible to detect rumors. Their main metric is the average number of tweets versus the number of followers and followees.

In [11] the authors analyze Twitter information flows during the 2011 Tunisian and Egyptian uprisings. Tweet senders are classified into "actor types", and the paper looks at how different "actor types" produce and pass information over Twitter.

The authors of [27] also analyze Twitter activity having to do with a severe thunderstorm at the 2010 Pukkelpop music festival in Hasselt, Belgium. The growth of tweets relating to the incident is documented, as well as the number of original tweets vs retweets. The paper finds that tweets from reliable sources received more retweets regardless of the content.

Evolutionary clustering and community detection approaches designed for dynamic social networks have been considered by [3], [7], [10], and [26]. The papers analyze the evolution of clustering obtained by using a fixed temporal smoothness penalty to the cost function of a static clustering method. These were applied to (a) the established k-means clustering problem, (b) a proposed evolutionary spectral clustering problem, and (c) FacetNet discovering communities relying on maximum a posteriori estimation based on observed networked data and a given prior distribution.

In [28], the authors consider an adaptive methodology for communities. They mention that statistical clustering is generally outperformed by evolutionary clustering since it produces clustering results aligning with long-term trends yet robust to short-term perturbations. Their algorithm adaptively estimates the smoothing parameter using shrinkage estimation, based on naïve estimate. The strength of their framework is that it extends a variety of static clustering algorithms, including hierarchical, k-means, and spectral clustering, into evolutionary clustering algorithms.

The authors of [22] discuss methods of extracting news from tweets. They use a Naïve Bayes classifier to separate news and non-news. The news tweets are partitioned into communities that posses similar characteristics using modularity measurements. In the current research, we a modularity based clustering into communities algorithm, different than the ones used in the references above.

3 Methodology

To study how Twitter is used in an emergency, posts that used the hashtag *#ParisAttacks* were collected for a 24-hour period. The terrorist attacks occurred from 21:20 to 21:53 on 15 November 2015 and the period for which data was collected covered 02:00 on 16 November to 02:00 on 17 November 2015. A network was created by representing each Twitter user as a node and directed edges to indicate a user referenced or mentioned another user in a post. we identified the following potential metrics to capture the network as it grew: k-Core (and the core is identified by the highest value of k before the $k+1$-core vanishes), number of communities, centrality, and betweenness.

The end-state Twitter network which tracked the use of *#ParisAttacks* grew to a final state of $86,821$ nodes and $105,601$ directed edges in the 24 hours we captured. The edges represent tweeted or mentioned users in the network. We thus obtain a directed network with weights on the arcs, whose relevant metrics of the end-state network are presented in Table 1.

Average Node Degree	2.433
Maximum Degree	2596
Average Path Length	6.418
Diameter	25
Modularity	.9521
Average Clustering Coefficient	0.084
Number of Communities	6,829
The Core	a k-core for $k = 13$

Table 1. Ground Truth *#ParisAttacks* Network's Attributes

The network data was divided into two-hour segments of time and reviewed for the above characteristics to monitor change as the network grew to the state shown in Table 1. This temporal analysis for the network was completed by incrementally adding two hour segments of data gradually building a full network. The time slice period was chosen for ease in managing the data amounts and because of the hypothesis that the end-state network characteristics would become cemented early in the network's growth.

The method of study focuses on quantifying each time period at the number of nodes, edges, k-Core, centrality and communities in comparison to the end-state network. These attributes are then used to compare across all time increments. Communities and centrality are reviewed with a second parameter for comparison.

The top ten nodes with the highest betweenness centrality scores are recorded for each time period and compared to the end-state network's top ten. The betweeness centrality similarity as the network grows is the percent of nodes common to both networks compared. Due to the large number of communities in the network, the study focuses on the top six communities for each time period which account for 73% of the data. As communities align differently at each time-step, an adaptive method was needed: the largest six communities are therefore appraised for their network representation. In other words, what percentage of the total number of nodes in the network at the given time-stamps are members of the six largest communities. These characteristics will then be compared over time to observe possible changes as the network expands.

4 Results and Analysis

To keep track of the network's growth the number of nodes, edges, and communities were tracked. As Twitter users mention one or more other users, the rate of growth in number of edges and nodes were both nearly linear as shown in Figure 3.

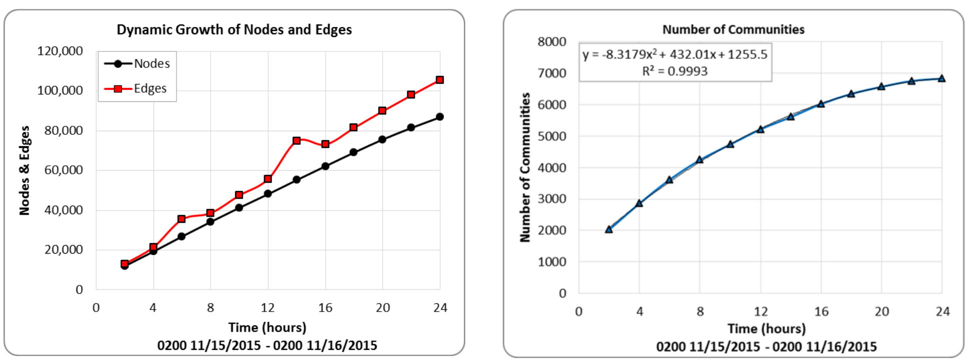


Fig. 3. Network Growth Over 24 Hours

We believe this is due to the events all happening in the beginning in a short amount of time, which constantly brings the attention to the event. On the other hand, if events would have been spread out over the 24 hour period, we would have expected exponential increases. This is supported by the slow increase in the number of communities, which began to plateau at the end of the 24-hours. It is best represented by a negative second order polynomial with an R-squared value of over 0.999 as shown in Figure 3.

The end-state network has a core identified by the 13-core, consisting of major news broadcasting stations: French President Hollande, French Newspaper,

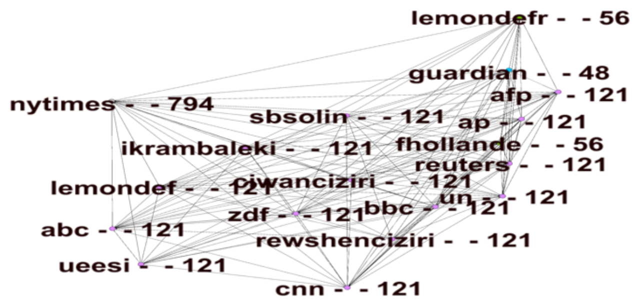


Fig. 4. End-state communities labels (numbers) in the Core of the Network

German Newspaper, ABC News, The Guardian, Reuters News, AFP News, BBC News, New York Times, United Nations, Time Magazine, CNN News and Associated Press. This 13-core was present within the first hour of collected Twitter feed and was constant throughout. This core contains the major news channels at each timestep, thus a very good measure to identify the key players driving the network during its inception.

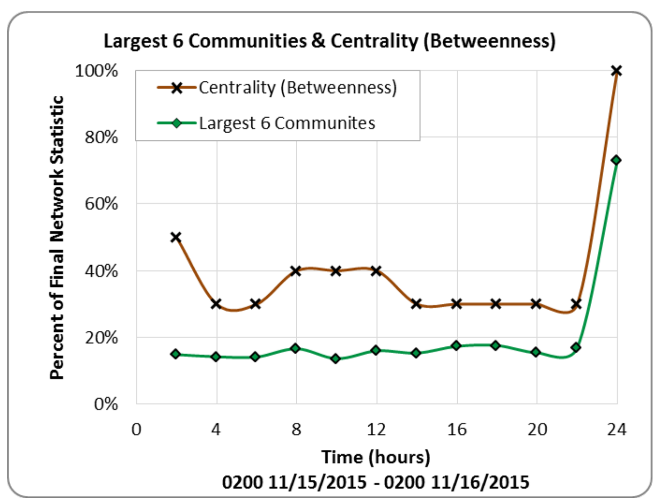
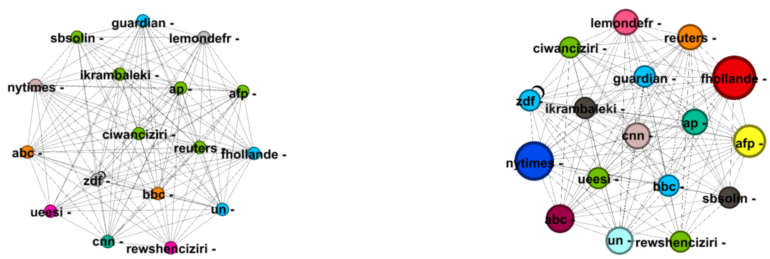


Fig. 5. End-state Network Community and Betweenness Centrality Overview

The communities were analyzed during the network's growth. The end-state network was partitioned into 6,829 communities, though 73% of the nodes were represented by only the six largest. Therefore, the six largest communities were tracked throughout the two-hours time slice networks, and compared to the end-state network. Of the $6,829$ Communities in the end-state network , the top six communities constitute of $28,966$ nodes (33.36%) $24,097$ nodes (27.75%) $5,006$ nodes (5.77%) $2,350$ nodes (2.71%) $1,463$ nodes (1.69%), and $1,092$ nodes (1.26%). The community labelled 121 is the largest in the network, and it contains most of the nodes of the core as shown in Figure 4.

Both centrality and the largest six communities were very unstable in nature during the growth of the network, but not the core (at $k = 13$ the whole time).

A representation of the network's largest six communities, and their comparison and the top ten most central nodes of the time slices were compared in Figure 5. The end-state network's closeness and betweenness centralities is shown in Figure 6.



(a) Closeness (b) Betweenness

Fig. 6. Nodes with High Centrality values in the End-state Network

Notice some overlap between the core at any stage in the 24 hours, and the nodes with high closeness and betweenness centrality. Namely AFP News (afp), CNN News (cnn), and French Pres Hollande (fhollande) are in the core and have high closenss centrality at the 24 hour mark. Similarly, AFP News (afp) Reuters News (reuters) and username ciwanciziri are in the core and have high betweenness centrality at the 24 hour mark. Therefore, the core in the early development of the network contains some of the representative main actors of this social network.

5 Conclusions and Further Studies

Analysis of the Twitter hashtag $\#ParisAttacks$ activity over 24-hours explored a network's topological metrics that were both descriptive and prescriptive of the end-state network. Knowledge of the end-state network and the associated metrics enabled a comparison of the dynamic network characteristics as it grew. The top central nodes have not been consistent, where only a couple of the top ten nodes in the end-state network have appeared in the top ten at earlier stages. Both the betweeness and communities experienced an unexplained spike at the end of the observed period. This spike would require additional research to discern if the rapid increase is a step function inherent to the network or an anomaly. This was not present in the 13-core.

The 13 k-core was representative of the end-state network within the first hour of Twitter activity, remaining consistent throughout the 24-hour period. This solidifies our belief in the stability and representative nature of the k-core metric that could be used for emergency Twitter activity.

6 Acknowledgments

The authors would like to thank the DoD for partially sponsoring the current research.

References

[1] Adam Acar and Yuya Muraki. Twitter for crisis communication: lessons learned from japan's tsunami disaster. International Journal of Web Based Communities, 7(3):392–402, 2011.

[2] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment, 2008(10):P10008, 2008.

[3] Yun Chi, Xiaodan Song, Dengyong Zhou, Koji Hino, and Belle L Tseng. On evolutionary spectral clustering. ACM Transactions on Knowledge Discovery from Data (TKDD), 3(4):17, 2009.

[4] Michele Coscia, Fosca Giannotti, and Dino Pedreschi. A classification for community discovery methods in complex networks. Statistical Analysis and Data Mining, 4(5):512–546, 2011.

[5] Peter Csermely, András London, Ling-Yun Wu, and Brian Uzzi. Structure and dynamics of core/periphery networks. Journal of Complex Networks, 1(2):93–123, 2013.

[6] Martin G Everett and Stephen P Borgatti. Analyzing clique overlap. Connections, 21(1):49–61, 1998.

[7] Tanja Falkowski, Jorg Bartelheimer, and Myra Spiliopoulou. Mining and visualizing the evolution of subgroups in social networks. In Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, pages 52–58. IEEE Computer Society, 2006.

[8] Anatoliy Gruzd. Netlytic. https://netlytic.org/home/, 2006–2016.

[9] Frank Harary, Robert Z Norman, and Dorwin Cartwright. Structural models, 1965.

[10] Yu-Ru Lin, Yun Chi, Shenghuo Zhu, Hari Sundaram, and Belle L Tseng. Analyzing communities and their evolutions in dynamic social networks. ACM Transactions on Knowledge Discovery from Data (TKDD), 3(2):8, 2009.

[11] G. Lotan, Erhardt Graeff, Mike Ananny, Devin Gaffney, Ian Pearce, et al. The arab spring— the revolutions were tweeted: Information flows during the 2011 tunisian and egyptian revolutions. International journal of communication, 5:31, 2011.

[12] R Duncan Luce and Albert D Perry. A method of matrix analysis of group structure. Psychometrika, 14(2):95–116, 1949.

[13] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. Twitter under crisis: Can we trust what we rt? In Proceedings of the first workshop on social media analytics, pages 71–79. ACM, 2010.

[14] Robert J Mokken. Cliques, clubs and clans. Quality & Quantity, 13(2):161–173, 1979.

[15] Mark EJ Newman. Modularity and community structure in networks. Proceedings of the National Academy of Sciences, 103(23):8577–8582, 2006.

[16] Gergely Palla, Albert-László Barabási, and Tamás Vicsek. Quantifying social group evolution. Nature, 446(7136):664–667, 2007.

[17] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. Nature, 435(7043):814–818, 2005.

[18] Samuel P Prahlow, Ralucca Gera, Henry Escuadro, Linda Eroh, and Steven Winters. Exploring k-dense graphs within graph theoretical analysis. 2014.

[19] Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi. Defining and identifying communities in networks. Proceedings of the National Academy of Sciences of the United States of America, 101(9):2658–2663, 2004.

[20] Nancy Roberts and Sean F. Everton. Timeline: What happened in paris attacks. https://sites.google.com/site/sfeverton18/research/appendix-1, November 2014.

[21] Kazumi Saito, Takeshi Yamada, and Kazuhiro Kazama. Extracting communities from complex networks by the k-dense method. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 91(11):3304–3311, 2008.

[22] Jagan Sankaranarayanan, Hanan Samet, Benjamin E Teitler, Michael D Lieberman, and Jon Sperling. Twitterstand: news in tweets. In Proceedings of the 17th acm sigspatial international conference on advances in geographic information systems, pages 42–51. ACM, 2009.

[23] Stephen B Seidman. Network structure and minimum degree. Social networks, 5(3):269–287, 1983.

[24] Stephen B Seidman and Brian L Foster. A graph-theoretic generalization of the clique concept*. Journal of Mathematical sociology, 6(1):139–154, 1978.

[25] Huawei Shen, Xueqi Cheng, Kai Cai, and Mao-Bin Hu. Detect overlapping and hierarchical community structure in networks. Physica A: Statistical Mechanics and its Applications, 388(8):1706–1712, 2009.

[26] Chayant Tantipathananandh, Tanya Berger-Wolf, and David Kempe. A framework for community identification in dynamic social networks. In Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 717–726. ACM, 2007.

[27] Teun Terpstra, A de Vries, R Stronkman, and GL Paradies. Towards a realtime Twitter analysis during crises for operational crisis management. Simon Fraser University, 2012.

[28] Kevin S Xu, Mark Kliger, and Alfred O Hero Iii. Adaptive evolutionary clustering. Data Mining and Knowledge Discovery, 28(2):304–336, 2014.