JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
ICASET
2024

Research Article
Collection: Intelligent Computing of Applied Sciences and Emerging Trends

Excavation Network Properties on Twitter Based Data of Bahauddin Zakriya University Multan

Yasir Arafat [1,2*], Alia Saeed[3], Amna Mukhtar[4], Mohsin Abbas[1,2], and Ahsan Saeed[1]

[1]Government College University Faisalabad Layyah Campus, Layyah, 31200, Pakistan.
[2]Khwaja Fareed University of Engineering & Information Technology, Rahim Yar Khan, 6400, Pakistan.
[3]The Islamia University of Bahwalpur , Bahwalpur, 63100, Pakistan.
[4]National University Of Modern Languages Islamabad, Islamabad, 44000, Pakistan.
*Corresponding Author: Yasir Arafat. Email: yaxirmalik@gmail.com

Academic Editor: Salman Qadri **Published:** April 01, 2024

Abstract: Social media plays a vital role in all aspects of our life. Critical issues of Bahauddin Zakriya University Multan have been disused in social media, involving datesheet, Results, Exams, Online exam and other students affairs and activities. This paper aims at the study for the detection of hidden communities and influential users on social media through analysis of tweets related to BZU. In this study, we discussed large scale analysis of twitter data using social network analysis techniques using different networks. We use methodology to analyze macroscopic and microscopic properties of network. Further, we detect communities and sub-communities and their impact on network. We also tried to extract information from these communities. We use 3 community detection algorithms and compare their results. As a result, we were able to visualize the communities and sub communities of large network. This approach may help in identifying positive and negative communities in a network.

Keywords: Social Network Analysis; BZU; Communities Detection.

1. Introduction

Bahauddin Zakriya University (BZU) is the largest Public University of South Punjab. It offers many undergraduate, postgraduate and Doctorate programs. Many other institutes are also affiliated with it. It provides affiliation of BA and B.Sc., MA and MSC, LLB and many other programs. Social media has become important part of our lives. We can get news of the any part of world within minutes. For analysis of any network we can get data from social media. Twitter is one of the social media platforms for information sharing. Twitter allows the sharing of short private or public messages which are known as tweets. People use twitter platform to interact with each other by sending retweet, mentions, hash tags. We can get data related to a certain network from twitter.

A network consists on nodes and edges. Node is a vertex and edge is link between two vertexes. A network has many nodes and edges. A network may be directed or undirected. The properties of a network are categorized in three categories that are microscopic properties, macroscopic properties and mezoscopic properties. Microscopic properties are at node level. They describe the out degree and in degree of a node. Out degree represent the outwards links and in degree represent inward links of a node. Average degree is the total degree of a graph. Macroscopic properties are at network level of a graph. These properties describe diameter, radius, average path length, distance, centrality and betweeness.

2. Materials and Methods

This paper consists on dataset on twitter based data of BZU. Data is extracted through API. Data consists of 7068 Jason files and each Jason files contains information about tweets. We use Gephi for computing microscopic and macroscopic properties. We also detect communities through Gephi. Different algorithms are implemented on Python to compute modularity of a network. We use different networks like mention, hash tags, retweet, in reply to user id to compute their properties and detecting their communities and sub-communities. We calculated microscopic and macroscopic properties to disclose the influencer user of networks. Then detect communities to know the hidden information of a network. Communities are the sub-structures of nodes in a network that are densely connected to their members and sparsely connected to other nodes

3. Results

There are the following macroscopic and microscopic properties that are calculated through Gephi
3.1 Microscopic and Macroscopic properties

The microscopic properties of the hash tag networks are as following.

- There are total 936 nodes and 1177 edges in this network.
- The Average degree of this network is 2.557.
- The Density of graph is 0.003.
- The Average Weighted Degree is 3.562.

It can be visualized as:

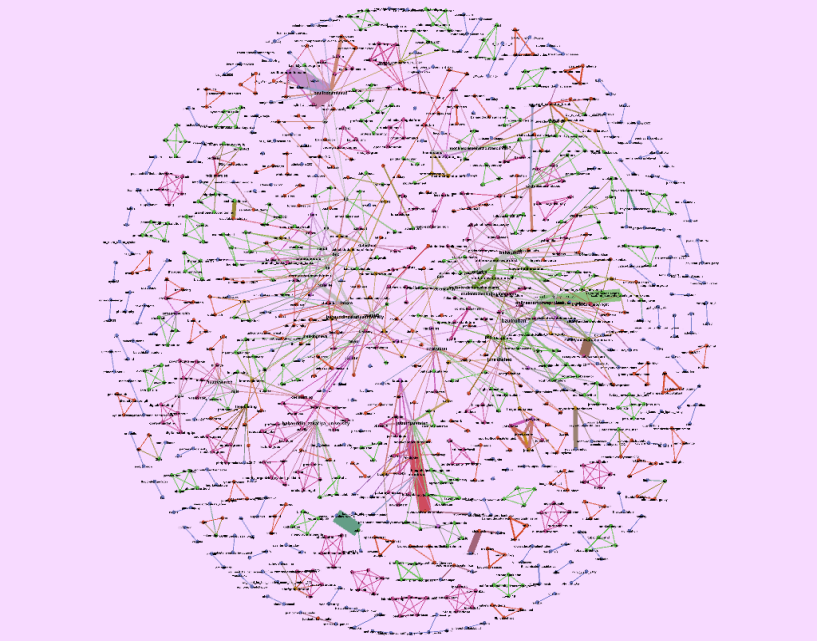

Figure 1. Visualizing the Network of Hash tags.

There are some high degree and low degree nodes in this network. Size of node is based on degree. Nodes that have high degree are larger in size. After calculating their properties, we analyzed top five to ten nodes. We analyzed why these nodes have high page rank, betweeness, closeness centrality etc. it is not necessary that all nodes that have high page rank, also possess high hits, closeness centrality or betweeness centrality. Some nodes may have high page rank and some may have high closeness centrality. It is because the page rank of any node could be increased by other nodes that are well-known or famous personalities.

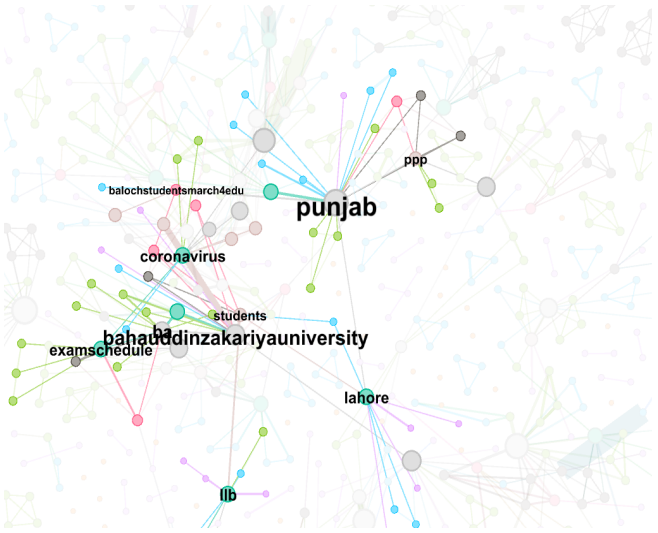

Figure 2. Betweenness Centrality

These nodes have high betweeness centrality. Coronavirus was spreading all over the world. Due to this, all institutes, offices, private and public companies were shut down. It was trending hash tag at that time. So, administration, faculty and students use hash tag coronavirus in their tweets widely. So, this hash tag has high Betweeness centrality. Exam schedule, students and BA are also related hash tags so they have high Betweeness centrality. Baloch student march was protest by the students of Baluchistan for their education. The students travel to Islamabad from Lahore. Hash tag Baloch student march was also used by the many peoples. As Maryam Nawaz visited the Baloch students march camp, students at BZU also support Baloch students so, this hash tag has high Betweeness centrality.

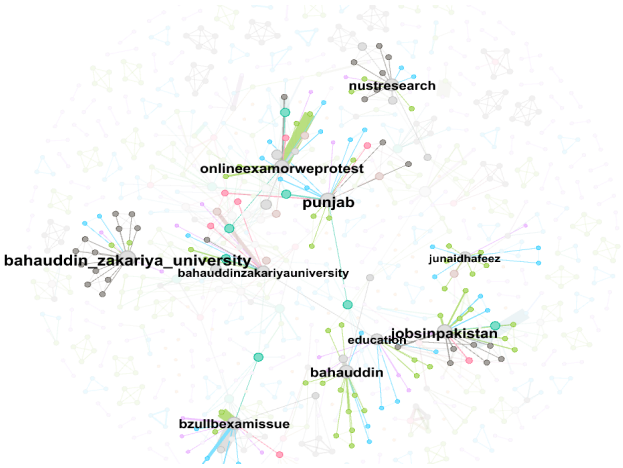

Figure 3. Highest page ranks

These are top nodes that have high page rank. Junaid Hafeez was the lecturer at BZU. On March 13, 2013 he was arrested byauthorities as he was accused by students to insulting Hazrat Muhammad (P.B.U.H) on social media. On December 2019, He was sentenced to death by District and session court Multan. So, it was trending hash tag that #freejunaidhafeez. That's why it has high page rank. Jobs advertisements of university use hash tag of jobs in Pakistan. So, its page rank is high. NUST research and innovation is also widely used # in the research area. So, many people use hash tag nust research and innovation. NUST collaborated with many institutes like QAU, KSU for the seminars and conferences related to technology. So, many institutes used hash tag nust research so its page rank is high. During COVID, online exams were scheduled across the country. So, hash tag online exams or we protest was trending because some people wanted online exams and some wanted physical exams.

3.2. Mezoscopic Properties

There are the mezoscopic properties of a network which are as following.

Communities and sub-communities Detection of Hash Tag Network

The modularity of this network obtained through Gephi is 0.586. Following is the modularity obtained through different algorithms.

1. Kernighan = 0.426511
2. Greedy = 0.808210
3. Louvain = 0.818845

There are 158 communities in hash tag network. Communities of this network can be visualized as:

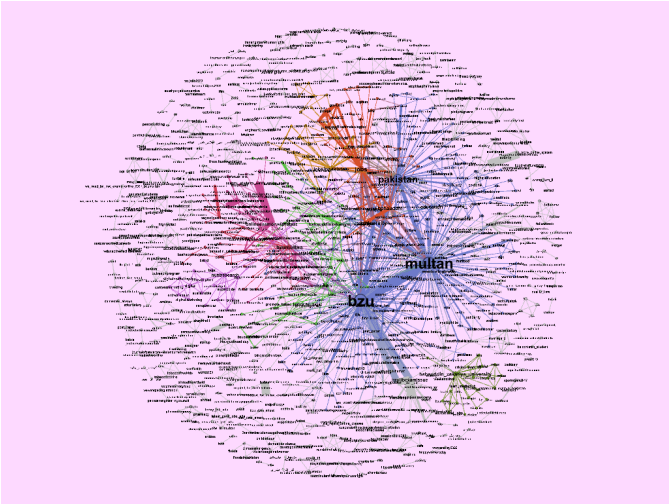

Figure 4. Communities of Hash Tag

BZU, Multan, Pakistan and Punjab are top communities of this network. As Bzu is the largest university in south Punjab and situated in Multan so, these hash tags are used frequently. Different colors show different communities some are large while some are small communities.

3.3 Sub-communities

Sub-communities are the communities in a community. They are represented by different colors.

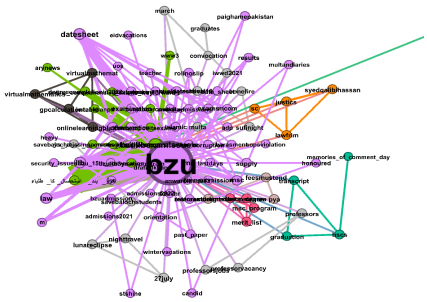

Figure 5. Sub-communities of BZU

Sub-community Bzu contains further communities. There are 105 nodes and 188 edges in this graph. Modularity of this graph is 0.418. There are 14 communities in this graph. Different colors represent different communities which are results, exams, transcripts, jobs, professors, graduated, winter vacation, past papers etc. These sub-communities are related to university.

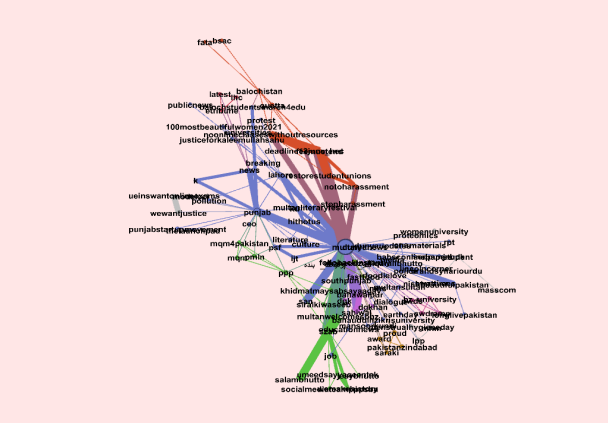

Figure 6. Sub-community of Multan

These are the sub-communities of Multan community. It has 101 nodes and 190 edges. Its modularity is 0.428. There are 13communities in it. These sub-communities are related to Multan and Punjab. Further sub-communities are Siraki, Women College, Bzu, and south Punjab, PPP etc. all these sub-communities of this graph. In south Punjab, there is majority of Siraki people. So, this is also sub-community of this network. Moreover, many nodes of one community are linked with other community.

4. Discussion

4.1 Small World Property Test

Real world networks are different from random networks. They have unique properties. Small world have high clustering coefficient and short average path length. While random network have low clustering coefficient and short average path length. Small world networks have high clustering coefficient because of their neighborhood.

First, we compute the avg path length and clustering coefficient of hash tag network

Then we generate a random model that consists on same no of nodes and edges as hash tag network

Then we calculate the clustering coefficient and average path length of that network.

We compare the avg path length and clustering coefficient of both networks and compare them.

5. Conclusions

Results of these algorithms show that Louvain is the fastest algorithm and it gives result better than other algorithms. Greedy and Louvain algorithms have minor differences so greedy algorithm gives also better results. From microscopic and macroscopic properties we disclose influencer users of the network. From communities we detect information of the network which cannot be detected by simple visualization. Results show that large communities are dominated by students of LLB, Dietitians Imran Khan supporters and physical exams protestors. Then, the avg path length and clustering coefficient computed through Gephi is compared. Results showed that the average path length of both networks real world and random network is approximately same but real world network have high clustering coefficient and random network have very low clustering coefficient. This showed that it is random network. Further, its graph on log-log scale is partially scale free graph. That's mean it is a random network.

References

1. Abbasi, A., J. Altmann, and L. Hossain, Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. Journal of informetrics, 2011. 5(4): p. 594-607.
2. Said, A., et al., Mining network-level properties of Twitter altmetrics data. Scientometrics, 2019. 120(1): p. 217-235.
3. Otte, E. and R. Rousseau, Social network analysis: a powerful strategy, also for the information sciences. Journal of information Science, 2002. 28(6): p. 441-453.
4. Wu, I.-C. and C.-Y. Wu, Using internal link and social network analysis to support searches in Wikipedia: A model and its evaluation. Journal of Information Science, 2011. 37(2): p. 189-207.
5. W. Xia, Y. Wen, C. H. Foh, D. Niyato, and H. Xie, "A survey on softwaredefined networking," IEEE Commununication Surveys & Tutorials, vol. 17, no. 1, pp. 27–51, 2015.
6. B. Wang, Y. Zheng, W. Lou, and Y. T. Hou, "DDoS attack protection in the era of cloud computing and software-defined networking," in 2014 IEEE 22nd International Conference on Network Protocols, pp. 624–629, IEEE, 2014.
7. F. Hu, Q. Hao, and K. Bao, "A survey on software-defined network and openflow: From concept to implementation," IEEE Communications Surveys & Tutorials, vol. 16, no. 4, pp. 2181–2206, 2014.
8. V. Pacifici, F. Lehrieder, and G. Dán, "Cache capacity allocation for bittorrent-like systems to minimize inter-isp traffic," in 2012 Proceedings IEEE INFOCOM, pp. 1512–1520, 2012.
9. Y. Huang, T. Z. J. Fu, D. M. Chiu, J. C. Lui, and C. Huang, "Challenges, design and analysis of a large-scale P2P-VoD system," in Proceedings of the ACM SIGCOMM 2008 conference on Data communication, vol. 38, pp. 375– 388, 08 2008.
10. S. Borst, V. Gupta, and A. Walid, "Distributed caching algorithms for content distribution networks," in Proceedings 29th IEEE International Conference on Computer Communications (INFOCOM 2010, San Diego CA, USA, March 15-19, 2010), (United States), pp. 1–9, Institute of Electrical and Electronics Engineers, 2010.
11. M. Qian, Y. Wang, Y. Zhou, L. Tian, and J. Shi, "A super base station based centralized network architecture for 5g mobile communication systems," Digital Communications and Networks, vol. 54, 03 2015.
12. S. Sharif, M. H. Y. Moghaddam, and S. A. H. Seno, "Adaptive cache content placement for software-defined internet of things," Future Generation Computer Systems, vol. 136, pp. 34–48, 2022.
13. W. Rafique, A. S. Hafid, and S. Cherkaoui, "Softcaching: A framework for caching node selection and routing in software-defined information centric internet of things," Computer Networks, vol. 235, p. 109966, 2023.
14. Y. Cui, J. Song, M. Li, Q. Ren, Y. Zhang, and X. Cai, "Sdn-based big data caching in isp networks," IEEE Transactions on Big Data, vol. 4, no. 3, pp. 356–367, 2018.
15. D. Kim and Y. Kim, "Enhancing ndn feasibility via dedicated routing and caching," Computer Networks, vol. 126, 07 2017.
16. E. Aubry, T. Silverston, and I. Chrisment, "Srsc: Sdn-based routing scheme for ccn," in Proceedings of the 2015 1st IEEE Conference on Network Softwarization (NetSoft), pp. 1–5, 2015.
17. J. Badshah, M. Kamran, N. Shah, and S. A. Abid, "An improved method to deploy cache servers in software defined network-based information centric networking for big data," Journal of Grid Computing, vol. 17, 06 2019.