

---

# Social Networks Analysis using Compressive Sensing

---

Anonymous Author(s)

Affiliation

Address

email

## Abstract

Social Networks are well-known for their massive scale. Analyzing the social networks to extract meaningful information is one of the most significant aspects of managing such massive user data. Larger the data, more valuable is the extracted information. But such scale of data is not always an advantage. When social networks have billions of nodes and relationships, the traditional graph algorithms that run in polynomial time pose a significant disadvantage for analyzing the social networks. Therefore, in this paper, we will try to find ways to apply compressive sensing to social networks. The idea is to find some set of sparse characteristics in a social network that allows us to group the users together and makes it easier to analyze the entire social networks in groups.

**Index** - compressive sensing, social networks, inter-community detection.

## 1 Introduction

Graphs have always been an efficient way of representing large amount of data. The nodes stand for individual entities and the edges represent some non-trivial relation between them. This idea has been extended to represent billions of users and their mutual relationships in the real-world systems such as social networks [7]. A typical representation of social networks using graphs, or social graphs, is to represent the users as nodes and their relationships such as being friends, accessing similar content is represented by edges. The edges can further have some weights associated with them to denote the degree of associativity between the users. Such structural representation can often blow up in size. It becomes very difficult to look at all the nodes and edges to quickly retrieve any meaningful information. Something as simple as finding top  $k$ -nodes can take order of few minutes using polynomial time algorithms.

To avoid this problem, it's sometimes easier to group the nodes together and analyze the clusters of multiple nodes together. This shrinks down the problem of analyzing billions of nodes to analyzing a few hundred or thousand nodes depending on the degree of such clustering. Such groups, or *communities* of users correspond to the real-world social groups [2]. This makes such groups even more significant because of the important characteristics it stands for. These communities are the more intensely connected groups with relatively small number of edges between them. In this paper, we try to analyze applications of compressive sensing to identifying these small number of important edges that connect multiple communities. Naturally, by treating the graph as a signal, the problem gets converted into that of retrieving some  $k$ -sparse vector corresponding to  $k$  such edges [1].

The rest of the paper discusses brief background and current scenario of applications of compressive sensing to social networks in section 2. Then we take a closer look at one such algorithm called LSR-weighted in section 3. Then section 4 discusses various real-world applications to utilize such community detection. Section 5 describes the results of running LSR-weighted and shows how these communities look by 3D visualization. Then we conclude by looking at important characteristics and limitations of these algorithms and mention few possible enhancements.

## 38 2 Related Work

39 Within past decade, online social networks have emerged and have given a new dimension to complex  
40 graph networks. For example, according to Nielsen [2], worldwide users spend over 110 billion  
41 minutes on social media sites per month, which accounts for 22% of all the time spent online,  
42 surpassing even the time spent on email. In the analysis of social networks, the existence of missing  
43 data is almost inevitable because the aforementioned constraints may prevent access to entire data  
44 of the networks. Mostly, direct measurement of each individual node can be difficult, costly, and  
45 sometimes impossible due to massive scale, distributed management, and access limitation of real  
46 social networks.

47 The developments in compressive sensing started with the seminal works in [1]. The authors noted  
48 that the combination of  $l_1$ -minimization and random matrices can lead to efficient recovery of sparse  
49 vectors and also has strong potential to be used in many applications. For the last couple of years,  
50 CS has been considered in signal processing, but its role in network applications is still in its early  
51 stages due to some challenging issues [3]. In networks, measurements are restricted by network  
52 topological constraints which is again absent in existing CS research. In other words, for every  
53 measurement, only links that induce a path or connected sub-graph can be aggregated together in the  
54 same measurement [5]. As a result, compressive sensing for network applications is quite different  
55 from other CS problems, although it is interesting in its own right because we can represent many  
56 real-world systems by their graphs/networks.

57 The first paper we found that applied compressive sensing to large graph networks was [3] which used  
58 compressive sensing to computer networks to identify congested links that are vulnerable to packet  
59 drops. This research falls under the category of *network tomography* and states many theoretical  
60 constraints on using traditional compressive sensing techniques on graph networks. The algorithm  
61 proposed is quite efficient and serves the purpose of monitoring computer networks. But the main  
62 disadvantage it has is the biased output due to a technique called random walks. While randomly  
63 traversing the graph, the algorithm is biased towards nodes of high degree. So it fails to capture the  
64 weighted unbiasedness in a network.

65 This lack of weighted constraint is solved by considering weights while traversing the graph  
66 in [6]. This algorithm is also equally efficient and talks about important null-space property of  
67 measurement matrix which guarantees correctness on the algorithm. But this algorithm doesn't  
68 talk about the applications of Compressive Sensing in Social Networks. This topic is discussed in  
69 detail, probably for the first time, in [4]. This work describes how we can overcome the theoretical  
70 challenges discussed in [6] while enabling weighted unbiased output for detecting inter-community  
71 structures. The paper also briefly describes a relatively new  $l_1$ -minimization technique called Lasso  
72 [9]. This algorithm is very similar to LSR algorithm which is described below.

## 73 3 LSR Algorithm

## 74 4 Applications

75 With this algorithm, it's easy to identify the inter-community edges that define some clustered graph  
76 structure. Till now, we've seen that a graph can represent multiple types of relations. By interpreting  
77 the edges in various ways, we can use this same algorithm to group the users by different heuristics  
78 and extract different types of information. We discuss a few possible application here.

### 79 4.1 Friend Recommendations

80 The relationship between two users can be that of friendship. This relation itself can be described  
81 using two different entities of weighted and unweighted edges. If we define friendship relation exists  
82 if two users are online friends, it's unweighted relation. If we describe the degree to which they've  
83 mutual online friends, then it becomes a weighted graph with weight of the edge describing the degree  
84 of connectivity. By running LSR-weighted on such graph, we can easily group the users together.  
85 Such communities can then be used to suggest friends within the same community or across different  
86 communities.

## 87 4.2 Advertising Content

88 By representing the edges as the degree to which users access similar content, we can group the  
89 users together based on accessed content correlation. If this content is specifically the advertisements  
90 that users found relevant, then we get an interesting clustering of users based on similarity of  
91 advertisements. It can help targeted advertising [10] where users from the same community can get  
92 similar advertisements to optimize the user experience and in turn increase revenue.

## 93 4.3 Detecting Filter Bubbles

94 Social networks have been under close scrutiny for filtering the content according to what the user  
95 wants rather than exposing variety of content. By using this data of accessed articles, videos, photos,  
96 etc., we can identify groups of people who are restricted to a small subset of the shared data. Using  
97 these clusters, we can further suggest data consisting of different view points than what the community  
98 is restricted to.

## 99 4.4 Detecting False Information

100 Suppose we identify the groups of users interested in similar type of content. If we rely on users to  
101 notify this false information, there is a good chance that users that disagree with valid information  
102 might report it as false information just to discourage that source. Instead, we can aggregate this user  
103 input across different communities and effectively identify the real false information sources.

104 There could be plenty more applications where analyzing a cluster of users can be more effective and  
105 efficient rather than looking at individual user. These are small subset of such applications that could  
106 be extended further as more sophisticated approach to solve the corresponding problems.

## 107 5 Results

### 108 5.0.1 Dataset

### 109 5.1 visualization

#### 110 5.1.1 graphs?

## 111 6 Conclusion

112 In this paper, we've looked at compressive sensing as an effective way of handling massive scale data  
113 in social networks. We discussed the major theoretical limitations faced by compressive sensing in  
114 social networks analysis. Along with LSR-weighted algorithm, we also looked at various approaches  
115 studied so far. The implementation of LSR-weighted also produces some encouraging results seen  
116 in the results section after running the algorithm on real-world dataset from reddit [?]. The main  
117 limitation we've found here is the coherence in social network. When we've a sparse network, the  
118 measurement matrix can be empty where we fail to output any sparse edges. But overall, it works  
119 well and there are some interesting applications as well to utilize this information of clustered users.

## 120 7 Acknowledgements and Credits

121 We would like to thank Prof. Peter Chin for helping us understand the problem statement and giving  
122 a few basic ideas to begin investigation. The techniques taught during CS591C2 course were also  
123 helpful during this project.

124 For most of the project, we worked together as a group. A rough distribution of work can be seen as  
125 follows,

126 Sahil - initial research involving going through a few papers, help with coding, applications of  
127 LSR-weighted

128 ..

129 ..

130

## References

- [1] D. Donoho, "Compressed sensing," IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [2] Nielsen statistics and measurements, june 2010. [Online]. mobile/social- Available: <http://blog.nielsen.com/nielsenwire/online-media-accounts-for-22-percent-of-time-online>.
- [3] J. Haupt, W. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," IEEE Signal Processing Magazine, vol. 52, no. 2, pp. 92–101, Mar. 2008.
- [4] H. Mahyar, H. R. Rabiee, Z. S. Hashemifar, and P. Siyari, "UCS-WN: An Unbiased Compressive Sensing Framework for Weighted Networks," in Conference on Information Sciences and Systems, CISS 2013, Baltimore, USA, Mar. 2013.
- [5] M. Wang, W. Xu, E. Mallada, and A. Tang, "Sparse recovery with graph constraints: Fundamental limits and measurement construction," in IEEE INFOCOM, Mar. 2012, pp. 1871–1879.
- [6] W. Xu, E. Mallada, and A. Tang, "Compressive sensing over graphs," in IEEE INFOCOM, Apr. 2011, pp. 2087–2095.
- [7] Reddit - [www.reddit.com](http://www.reddit.com)
- [8] H. Mahyar, H. R. Rabiee, A. Movaghar, E. Ghalebi, and A. Nazemian, "CS-ComDet: A Compressive Sensing Approach for Inter-Community Detection
- [9] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society B, vol. 58, pp. 267–288, 1994.
- [10] [https://en.wikipedia.org/wiki/Targeted\\_advertising](https://en.wikipedia.org/wiki/Targeted_advertising)