Analyzing Funding Network in Turkey

Ufuk Ozkanli
Zirve University
Computer Engineering
Gaziantep, Turkey 27260
ufuk.ozkanli@zirve.edu.tr

Suat Mercan
Zirve University
Computer Engineering
Gaziantep, Turkey 27260
suat.mercan@zirve.edu.tr

Mehmet Hadi Gunes
University of Nevada - Reno
Computer Science and Engineering
Reno, NV 89557
mgunes@cse.unr.edu

***Abstract*—Funding agencies in many countries distribute an important amount of grant to institutes, universities and research companies. Collaboration network among institutions, research fields and researchers may exhibit interesting characteristics, and analyzing these networks may uncover important information which might help to set better policies. In this paper, we are analyzing funding network in Turkey. We consider researchers, departments and organizations as nodes and being included in the same project as relationship between entities. We investigate researcher network, organization network and department network by using social network analysis tools and techniques in order to understand interaction in terms of joint projects.**

***Index Terms*—social network analysis; funding network; complex networks**

I. Introduction

Analysis of social networks may reveal some hidden, but significant information. The interactions between entities can be understood better by examining the relationship among them as a complex network from local and global perspective. Social network analysis techniques and parameters characterizing communities or nodes give us the opportunity to obtain important knowledge about examined topic, community and structure. People recently tend to investigate different networks such as biological, transportation and social networks so that they will have a better understanding and possibly better solutions and improvements.

In this work, we are analyzing network of researchers, organizations and departments which are supported by The Scientific and Technological Research Council of Turkey (TUBITAK) [1] via a funded project. Project evaluation reports written by investigators are released by the TUBITAK after the projects are complete. So, we only include completed projects in our evaluations. In the reports submitted by investigators, much information such as researchers, researchers' organization and departments, keywords about the project topic, publications from project are provided. We focus on the structure of relationship in researcher, department and organization level.

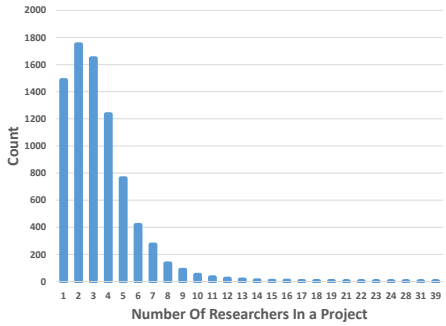
Similar studies that look for scientific collaboration among researchers have been done so far. Generally, people are interested in coauthorship of a paper or citations from one paper to others as a directed graph in order to see the connection. [2] is investigating collaborations in specific fields such as Computer Science, Physics and Biomedical. He considers two people connected if their name appear together on a paper. A recent report [3] is interested in that if geographical location is still effective in collaboration among scientist after the Internet removed the borders for knowledge. They look for the probability of a link existence between two cities in the collaboration network based on geographical distance. The closest work to our work is done by [4]. They present collaboration pattern among scientists by looking at funded projects by National Science Foundation (NSF). In our work, we are examining the case of Turkey which means the dataset is different. So, we expect different parameters which might be caused by grant policies imposed by governments, the way that scientists collaborate or some other reasons. We also guess that these parameters affect the output of the grants, but we don't focus on the comparison and mentioned reasons and outputs for now. We further look at department network which gives an idea about collaboration patterns among research fields.

In order to obtain project information, we had to develop a script which pulls all projects. A list of project is not given, we queried database by using category field which is assigned one to each project. By using this category field, we are able access all projects uniquely. There are 8089 completed projects between 1976-2012 and 18390 people worked in these projects. We apply some filtering in data explained in detail later in the next section.

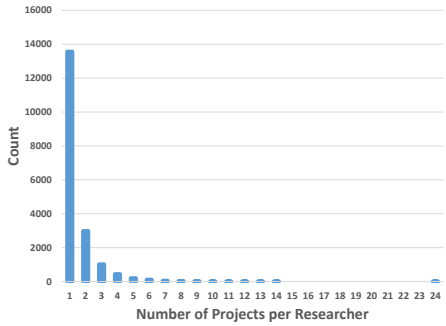
In the rest of paper, we first look at researcher network in Section 2. We analyze some statistical information in addition to complex network characteristics such as degree distribution, clustering coefficient, diameter etc. We also look at giant component if it is really different than the whole network. Section 3 does a deep investigation on organization network. Then we look at department network in Section 4 which gives important information about the relationship among collaborating fields. Finally the conclusion is driven in Section 5.

II. Researcher Network

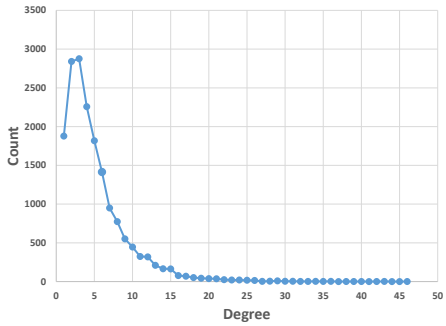
In this part, we are investigating the relationships among researchers. The way we construct the researcher network is as follows; we consider there is a link among people who worked in the same project. In original database18390 people and 8089 project exist. Some of people only did one project by himself (600), zero degree. We don't include these people into our network because they don't have any collaboration



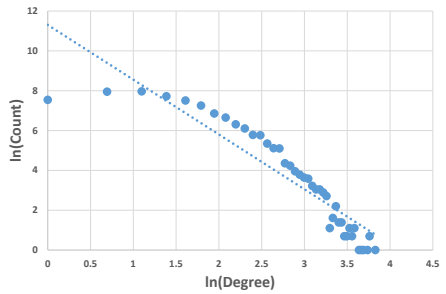
a. Number of projects vs. group size.



b. Number of people vs. number of projects they participate.



c. Degree Distribution of Researcher Network.



d. Log-Log Scale of Degree Distribution of Researcher Network

Fig. 1. Researcher Network Statistics and Degree Distributions.

with anybody. There are also very few projects to which a high number of people joined like 39 people in a project. We wanted also excluding these projects because it does seem realistic to us. It might also falsify peak values of parameters and suppress real interesting cases to appear. We excluded projects performed with more than 16 people (13 projects and 200 people). Eventually we have 8076 projects, 17449 researches and 44735 edges among researchers.

Figure 1.a gives the histogram for research group size. Horizontal axis represents number of people in a project and vertical axis tells the count. Some of the projects are individual while projects of a couple has the highest number. Almost 90 % of the projects are performed with at most five people. There are also very few unusual cases such that 39 people worked in one project.

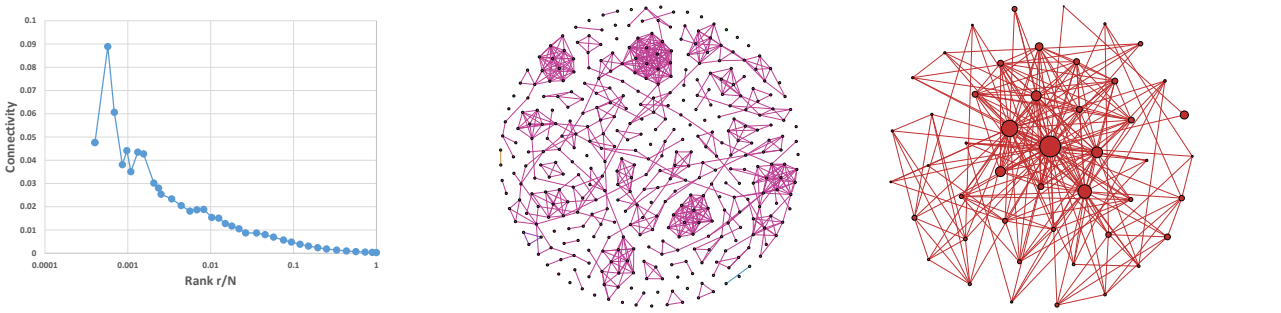
Figure 1.b gives information about number of projects that researchers participate. Most of the people (almost 14000, 80%) participated only one project. A considerable number of people joined to 2 and 3 projects. These people are hubs for connections between other people. A person takes role in 24 different projects but, he doesn't have highest degree. His degree is 16 which means that he has many individual projects.

When we analyze the researchers as complex network, there are 1748 components, one of which is giant that includes around 56 % of all nodes. This is usually what happens in social networks. The remaining components are small size. These small components are probably project groups whose members are less likely collaborating with others in other projects.

Figure 1.c shows degree distribution of researcher network. Average degree is 5.1 and weighted average degree is 5.6. The highest degree is 46 and there is a considerable number of researcher with degree 1 and 2. Figure 1.d shows degree distribution in log-log scale. The data is close to a power law distribution especially in the tail. Also, it shows similar behavior with other similar other co-authorship networks.

Clustering coefficient is 0.87 which is really high. Clustering coefficient of more than half of the nodes is 1 since these people take role only in one project and all people in one project are linked to each other. The diameter of researcher network is 33 and average path length is 11.3. High clustering coefficient and small path length are typical properties of small-world networks.

When we look at rich club phenomenon, we find the assortativity of researcher network as 0.33 which means that the network is assortative as defined in [5]. That is, nodes with high degree tend to cluster. In Figure 2.a, connectivity of the nodes is shown. Maximum rich club connectivity 0.09



a. Rich Club Connectivity. b. Researchers with High Degrees. c. Ego Network of Highest Degree Node.

Fig. 2. Rich Club Phenomenon in Researcher Network.

is obtained when top 0.05% of nodes is considered and for top 1% of the nodes, the connectivity is around 0.02. The connectivity is decreasing when we include more nodes with lower degree. We are getting a sparser network. Rich club is a typical property in scientific collaboration networks in which elite people tend to form groups as we see in our case. In Figure 2.b, we show nodes whose degree is higher than 17 in order to visualize the connections better. In this graph, there are 383 nodes and 769 edges. There are interesting motifs among different number of nodes and densely connected clusters among rich nodes. We also observe some thick edges which indicate frequent collaborations between same people. Ego network of the node with highest degree is shown In Figure 2.c. This node has connections with high degree nodes such as 36, 30 as well as with some other nodes with lower degree.

When we analyze the giant component separately, it is consisting of 9725 (56%) nodes and 31389 (70%) edges. Average degree is 6.4 and weighted average degree is 7.1. Degree is higher than that of global network as we expected since many people are participating more than one project in this group. Network diameter is 33 and average path length is almost same with the whole network. So small islands haven't affected the network properties much.

III. Organization Network

In this part, we are investigating the collaboration network among universities, institutes and some companies as shown in Figure 3.a. There are 289 different organization, but only 140 of them have collaboration. We see that most of the projects are done within organizations. So, there are 140 nodes, 324 edges and 648 collaborations (8% of projects) among organizations. Average degree is 4.6 and weighted average degree is 18, the difference is much since most organizations have multiple joint projects.

The distribution of degrees here follows a heavy-tailed behavior in Figure 3.b. Interestingly, even though Bozok University is not one of the popular university in Turkey and it is a pretty new founded university, it has the highest degree and highest number of projects. It has 250 collaboration with 29 different universities. Its betweenness parameter is also the highest one. Number of projects in an organization generally comply with the degree of organization when we look at the ranking table. But, there are some organization couples which don't appear in degree and project count rankings in top but they have a strong relation. Also, degree and weighted degree don't a direct relation.

Almost all the universities are connected via collaboration relationship in the giant component. Clustering coefficient is 0.65 which indicates that organizations have cliques but are also open to work with different organizations. Diameter is 7 and average path length is 3.3 which are low as expected in a small-sized network. But, we can still mention a small-world and a dense connection among nodes.

Assortativity of organization network is - 0.17 which means that rich nodes have connections with each other as well as with the nodes with small degrees. In Figure 2.c, connectivity of rich nodes is shown. Connectivity of top 10% (14) organization is 0.35 which shows a dense connection among rich nodes, but the connectivity is not too low for non-rich nodes. Bozok University and Hacettepe University are from top organizations in terms of project count and degree, this couple also holds top position in number of collaboration. Some small cliques and clusters are also being observed in the network (Figure 3.a).

IV. Department Network

In this part, we are investigating the relationships among departments in order to see the interaction among different disciplines. There are 394 different deparments, but people mostly do intradisciplinary projects rather than interdisciplinary. 78 different departments have 390 total and 93 distinct collaborations. Only 5% of all projects are interdisciplinary. So, we get a pretty small-sized network.

Biology department appears in 2706 projects and Chemistry department appears in 1319 different projects. The ranking of departments based on number of projects, degree and collaboration can be seen in Table I which gives some interesting

Organization	Project Count	Organization	Degree	Organization1	Organization2	Collaboration
Bozok Univ.	1714	Bozok Univ.	29	Bozok Univ.	Hacettepe Univ.	50
METU	1510	Istanbul Univ.	18	TUBITAK MAM	Sismologie Experimantale	28
Ege Univ.	1261	Hacettep Univ.	17	Bozok Univ.	METU	25
Hacettepe Univ.	1261	Ankara Univ.	17	Akdeniz Univ.	Laboratorie De.	24
Ankara Univ.	1155	Ege Univ.	17	Bozok Univ.	Ankara Univ.	23
ITU	986	Sabanc Univ.	17	Bogazici Univ.	Izmir High Tech. Inst.	22
Cukurova Univ.	622	ITU	15	Geology Foundation	Department of Mining	22
Dokuz Eylul Univ.	590	METU	15	Geology Foundation	Nature Protection	22
Gazi Univ.	563	Dokuz Eylul Univ.	15	SU Univ.	Water Products Inst.	21
Istanbul Univ.	521	Cukurova Univ.	13	Bilkent Univ.	NASU	20
Department	**Project Count**	**Department**	**Degree**	**Department1**	**Department2**	**Collaboration**
Biology	2706	Biology	34	Biology	Chemistry	53
Chemistry	1319	Geology Eng.	9	Environmental Eng.	Geology Eng.	13
Food Eng.	557	Chemistry	8	Biology	Mechanical Eng.	13
Chemistry Eng.	464	Environmental Eng.	7	Geophysics Eng.	Geology Eng.	12
Physics	460	Biochemistry	6	Biology	Medical Biology	11
Environmental Eng.	427	Earth Sciences	5	Biochemistry	Parasitology	10
Geology Eng.	421	Civil Eng.	4	Biology	Pediatric	9
Civil Eng.	362	Chemistry Eng.	4	Biology	Neurology	9
Botanic	320	Electrical Eng.	4	Physics	Geology Eng.	8
Mechanical Eng.	285	Computer Eng.	3	Biology	Cardiology	8

TABLE I
DEPARTMENT RANKINGS

information. Highest number of collaborations(53) is observed between Biology and Chemistry departments, two rich club. Biology appears 6 times in 10 of collaboration numbers. It has connection with many fields and interestingly with Mechanical Engineering. Geology has the second highest degree even though it is not in top ranking in project count. Geology has collaboration with Physics, Geophysics and Environmental Engineering.

Biology department has the highest degree (34) and weighted degree (215), it also has the highest betweenness parameter. Average degree is 2.3 and average weighted degree is 10. In degree distribution in Figure 3.d, there are a couple of nodes with high degrees and most of them has lower degrees. The diameter of the network is 7 and average path length is 3.

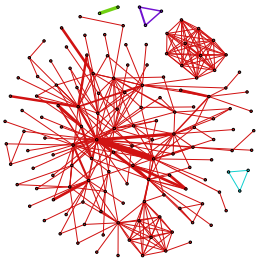
Assortativity coefficient is calculated as - 0.30 which shows a collaboration tendency between high degree nodes and low degree nodes. Even though two richest node (Biology and Chemistry) has a strong relation (thickest edge) which indicates something like club of richests. On the other hand, Biology department has connection with almost half of the nodes while lower degree nodes don't have much connections with each other. When we look at department network, Figure 3.d, in deep, Biology department places itself to the center and forms a star-like topology.

V. CONCLUSIONS AND FUTURE WORK

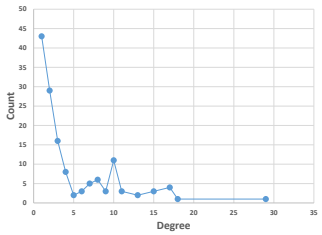
In this work, we have made a deep investigation about funding network in Turkey by using complex network analysis approach. We have gathered publicly available data from the website. We build three different networks; researcher network, organization and department networks from project information.

Degree distribution for all networks is not uniform. There are less number of nodes with high degrees and majority with lower degrees in general. This conforms to 80-20 rule which is observed in many social networks especially in scientific collaboration networks. We also look at degree distribution of researcher network in log-log scale and see that it resembles power-law especially in the tail. We also observe small-world feature where we have high clustering coefficient and small shortest path among nodes. We observe different rich club characteristics for networks. Researcher network has assortative mixing, organization network has non-assortative mixing and department networks have reverse assortative mixing. In department network, one of the most interesting findings is that Biology department collaborating with many other fields. But, we find total number of collaborations among different fields too less while the academia is celebrating the interdisciplinary research.

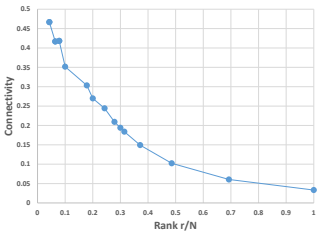
As future work; we have built researcher, organization and department networks in this work. Some more information such as keywords, publications form funded project etc. is provided in the website. We are planning to extend this work by including these additional information. We believe that this will give us more intuition about collaborations and beneficial statistical information. Moreover, funding of projects has gained momentum in recent years by economical development and government policy in Turkey. In this work, we only include completed projects. In the near future, we foresee that many new project information will be added to database. Analyzing a larger dataset will give a better idea about collaborations and an analysis in timeline might give an idea about the evolution. We also plan to make a comparison between our findings and similar studies performed for other countries especially for developed ones in order to see the differences which might help to improve and determine the



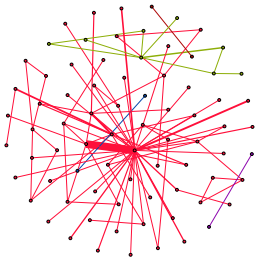
a. Organization Collaboration Network.



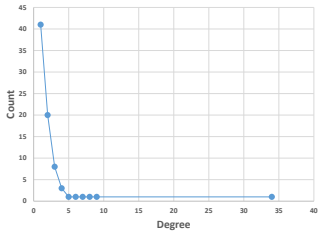
b. Degree Distribution.



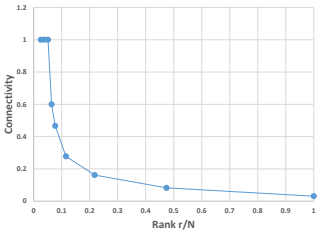
c. Rich Club Connectivity.



d. Departmental Collaboration Network.



e. Degree Distribution.



f. Rich Club Connectivity.

Fig. 3. Organization and Department Collaboration Networks.

policies about funding concept.

REFERENCES

[1] www.tubitak.gov.tr.
[2] Newman, Mark EJ. ”Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality.” Physical review E 64, no. 1 (2001): 016132.
[3] Raj Kumar Pan, Kimmo Kaski, and Santo Fortunato. ”World citation and collaboration networks: uncovering the role of geography in science.” Scientific reports 2 (2012).
[4] Hakan Kardes, Abdullah Sevincer, Mehmet Hadi Gunes, and Murat Yuksel. ”Six Degrees of Separation among US Researchers.” In Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on, pp. 654-659. IEEE, 2012.
[5] Newman, Mark EJ. ”Assortative mixing in networks.” Physical review letters 89, no. 20 (2002): 208701.