International Virtual Research Organization Network Evolution

Ashley B. Trudeau
University of Illinois at Urbana-Champaign
abt2@illinois.edu

Abstract

International Virtual Research Organizations (IVROs) are organizations established to foster collaboration between international groups using cyberinfrastructure, which provides mechanisms for organizing, planning, and executing scientific research. This study traces the evolution of the organizational network of a large multi-disciplinary IVRO over the course of nine years. Results show significant deviations in participation during certain years that may indicate organizational turbulence.

1. Introduction

In the last century, the research community has become more collaborative and intertwined than ever, particularly in STEM fields [1, 2, 3]. This trend has been amplified by the rampant developments of technology that has closed the distance between collaborators on different continents [4]. Because of this shift in research collaboration, it is important to understand the composition of collaborative research communities and their outputs. This analysis will uncover the evolution of the organizational network of a team-based International Virtual Research Organization focused on scientific work. Prior work addressing collaboration networks examined the connection between collaboration skills and performance of researchers [5], network analysis of information sciences [6], and collaboration network impact on publication [7].

Scientific teams are becoming more prominent in today's research landscape. By 2013, 90 percent of published scientific articles were products of collaborations between two or more authors [8]. The increase in collaborative research in recent years led to the development of a new field to learn how these teams operate and what makes them effective. Team science is interdependent, collaboratory research conducted by two or more people [8]. Teams may face a number of challenges including incompatible goals, geographic dispersion, and funding difficulties. The

members of research teams are typically also members of other organizations, groups, and teams, so they often have to prioritize the needs of the other collectives to which they belong.

An International Virtual Research Organization (IVRO) is a collaborative scientific research collective composed of several research centers from around the world. In order to study how various independent organizations come together to collaborate, it is important to understand the unique elements and structural dimensions of an inter-organizational relationship (IR). There are three basic elements of an IR. First, behaviors enacted by member organizations are directed towards a goal that is both collective and self-interested. Second, interdependence emerges as tasks and functions are divided amongst partner organizations. Finally, the IR has its own identity that is independent of its partner members.

IRs are marked by three major dimensions. First, collaborations are formalized in some way. According to Van de Ven [9], "an interagency agreement exists if any form of expression has been made between the parties regarding the terms of their relationship" (p. 26). Decision making processes in IR are unique from intra-organizational processes in that decisions are typically made by a committee or board of individuals who represent member organizations. Perhaps the most important dimension for this study is structural complexity of the IR. Structural complexity is defined by the number of different working parts in play and how they are incorporated into the functionality of an IR. An IR may become more complex with the introduction of more partner organizations, tasks, topics, or projects [8]. Similarly, research [10] also suggests that as more organizations enter the community, network density (the number of existing relationships divided by the number of possible relationships) tends to increase, making the community more complex. That being said, an organization cannot grow infinitely more complex without suffering a breaking point known as complexity catastrophe [10 and 11].

According to Bryant and Monge's [10] community evolution stage model, as an community of organizations changes, the types of ties in the network

HICSS

also change. The authors posit that, in the beginning stages of an organizational community, there are more mutually beneficial ties than competitive ties between organizations and overall network density grows rapidly. Then, as the organizations work to survive in the network, the number of competitive ties increases and mutual ties decrease while the overall density of the network increases at a slower pace. The community reaches the self-sufficiency stage when they are more equipped to survive despite changes in the environment. During this stage, mutual ties increase in proportion to competitive ties and overall density of the network reaches its highest point. When there are major, fundamental changes in the environment, the community may need to reorient in order to survive. During this period of time, both mutual and competitive ties decrease substantially.

This study will focus on how a project-based International Virtual Research Organization (IVRO) composed of several research organizations. The following research question will be explored:

> RQ1: How does a large, team-based IVRO network evolve throughout its lifespan?

2. Data

This dataset comes from a multi-year, IRB approved research project examining the organizational life of an international virtual research organization that began in 2010 with two research centers. In 2013, a third center was incorporated and by 2018, the IVRO was composed of seven partner research centers. The purpose of the organization was to engage scientists from various disciplinary and national backgrounds in addressing the pressing concerns of their fields. The structure of the organization is team-based. Every 6-9 months the organization holds a workshop, during which members come together to present accomplishments, ideas, etc. Workshops also give the attendees a chance to join an existing team or create a new team focused on a particular problem or topic.

Team progress is updated in annual reports released every summer. The network data was drawn from lists of projects in these annual reports from the beginning of the organization in 2010 until the last report in 2018. The organization is made up of professors, research scientists, post-docs, and graduate students from different places around the world and these members often participate in many different projects throughout the years. One of the major goals of the organization is to foster productive collaborations between scientists at all career stages.

The dataset in question was selected based on the organizational leaders' desires to see the progress of their initiative as well as how membership changed throughout the years.

3. Methods

In Social Network Analysis (SNA), relational data establishes connections between individual agents in a network [12]. In this study, there are two components to the network graph. First, nodes are the agents in a network who are connected to one another. Second, edges are the links that indicate a relationship between nodes [13]. The dataset captured the project membership data from 2010-2018, which shows connections between members based on common participation in at least one project that year. Therefore, nodes will be people (see Table 1) and edges will be relationships between them, either in common collaboration groups (see Table 1). The network is represented visually in a graph, which displays nodes and edges. A key characteristic of these graphs is the number of components it has, which indicate clusters of nodes that are connected to one another [12]. For example, when all nodes in the network are connected in some way to the large mass of nodes, there is one component in the network (such as the 2011 network in Figure 1), when there are two independent clusters of nodes that are not connected to one another (such as the 2012 network in Figure 1), there are two components in the network, and so on. The second type of data utilized in this study is attribute data, which represent various properties or characteristics of the nodes [12]. The primary attribute considered in this study was the participant's position in the organization. Three types of positions were distinguished: permanent members (Professors, research scientists, or industry professionals), post-docs, and graduate students. It is important to note the differences in status because the major goals of the organizations are to innovate as well as improve the skills of junior participants.

There was a significant change in the network throughout the years. Between 2010 and 2014, membership consistently grew but there was a definite change in the year 2015. Membership reached an all-time high in 2014 at 95 members, with one connected component. In 2015, this dropped to 56 members, with 3 components. The year 2015 seems to represent a transition period. In 2016, participation nearly doubled at 100 members.

This analysis will utilize several network metrics in order to address the research question. First, the network topology for each year will be examined in order to understand the nature of the connections being
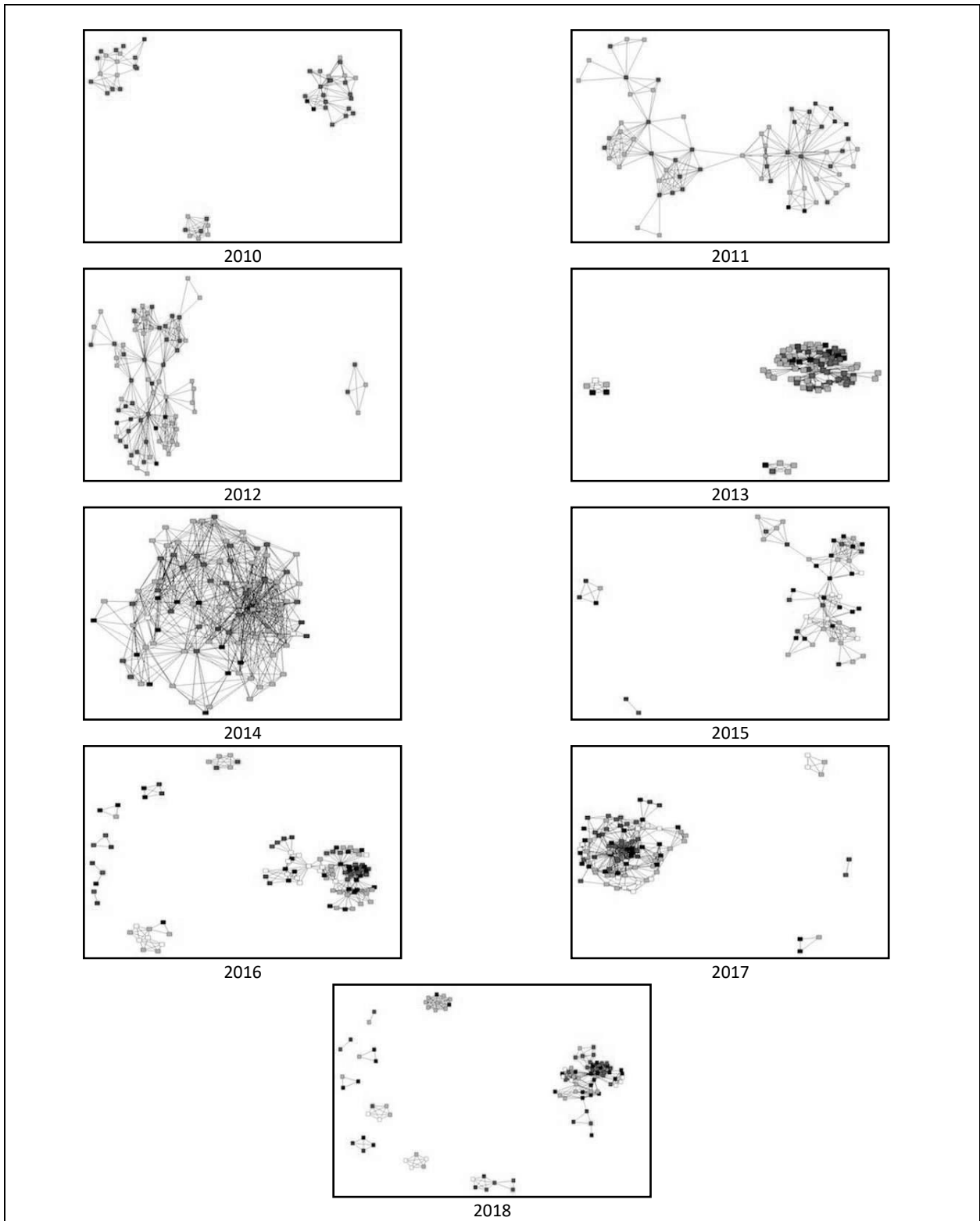
2010

2011

2012

2013

2014

2015

2016

2017

2018

Figure 1: Network Progression from 2010 - 2018

Table 1. Network Information

	2010	2011	2012	2013	2014	2015	2016	2017	2018
Nodes	54	66	76	90	95	56	100	96	95
Edges	157	234	302	352	647	183	366	506	371
Number of components	3	1	2	3	1	3	8	4	10
Degree Centralization	0.317	0.427	0.426	0.462	0.309	0.216	0.296	0.392	0.317
Density	0.110	0.109	0.106	0.088	0.145	0.119	0.074	0.111	0.083
Average distance	1.943	3.213	2.428	2.324	2.172	3.217	2.664	2.364	0.657
Clustering coefficient	0.911	0.898	0.916	0.917	0.541	0.859	0.899	0.680	0.929
Average Path Length	1.9	3.2	2.4	2.3	2.2	3.2	2.7	2.4	2.2

formed at the network level. Network topology refers to the "connection pattern" in a network [14]. There are three types of network topologies identified in this study. First, a small world network is one in which most nodes are not connected to all other nodes but can be reached within a small number of linkages. These networks are characterized by a high clustering coefficient and low average path length [14]. The clustering coefficient is a measure of how tight-knit or cliquish a network is. The range of possible clustering coefficients is between 0 and 1, where cliques with a coefficient closer to 1 are tightly clustered [14]. Average path length indicates the average number of links that separates a given pair of nodes [14]. Random networks are those in which each node has an equal probability of connecting with each other node in the network [14]. Random networks tend to have a normal distribution of node degrees. Degree centrality indicates the social influence or importance of a node in a network [14]. Therefore, the degree distribution shows how patterns in a network form based on the social influence of its members. Finally, preferential attachment refers to the idea that when faced with the option to connect with a person who has many connections versus a person who has few connections, one is more likely to establish a link with the well-connected person [15]. Preferential attachment networks tend to have a skewed distribution of node degrees. The resulting graphs show an unexpected progression throughout the years. Then, the degree centrality for each node will be calculated in order to find the most central participants in each year [14].

4. Analysis and Results

This analysis begins with a discussion of network level results. The years 2010-2013 reflect a relatively stable density progression with about 0.1 average

density (Figure 2). The emergence stage of the IVRO deviates from Bryant and Monge's [10] model in that the network density did not rapidly increase during these years. It is possible that the organization shifted to the maintenance stage during these years despite the lack of a significant increase in density. Unlike Bryant and Monge's [10] model, there is no indication of competitive ties in this network as relationships appear to be mutually beneficial.
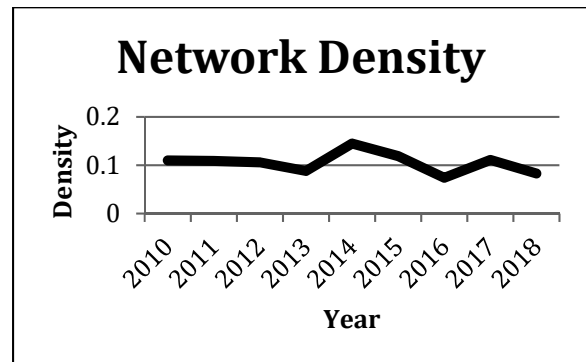

Figure 2: Network Density

In 2014, however, there was a massive shift. Network density spikes to about 0.145 in 2014 with a steep drop in the clustering coefficient. This shift seems to be indicative of the self-sufficiency stage [10], where mutual ties become enhanced and density reaches an all-time high. Visually, the graph (Figure 1) looks quite disordered when compared to previous years. Before the network data was assembled, it was expected that the 2015 graph would show the most contrasting result, but it seems that the organization grew much more complex in 2014, perhaps providing a window into possible mounting troubles.

Despite 2014 representing a period of self-sufficiency, the organization was not able to maintain the high density of the previous year and therefore

entered a period of transformation with a density of 0.119 in 2015 and reaching an all-time low of 0.074 in 2016. In 2017, network density rose again to 0.111 before dipping slightly to 0.083 in 2018. Though not entirely consistent with Bryant and Monge's [10] model, these two years appear to indicate a period of improvement and readjustment in response to the major shifts in 2015.

In order to understand network topology, node degrees were calculated and then plotted using a histogram (Figures 3-11). The 2010 network has a high clustering coefficient at 0.911 and a short average path length of 1.9. The distribution of node degrees (Figure 3) does not resemble either a normal or a skewed distribution.
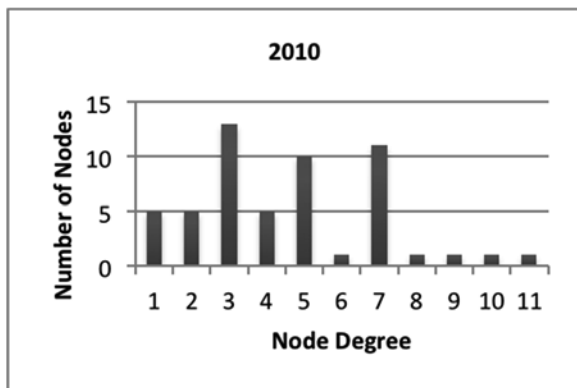
Figure 3: 2010 Degree Distribution

The 2010 network is a small world network due to its high clustering coefficient and short path length. The 2011-2013 (Figures 4-6) networks do appear to have slightly skewed distributions, but they also have very high clustering coefficients (2011: 0.898, 2012: 0.916, 2013: 0.917). These networks are small world networks with a slight lean towards preferential attachment networks due to their skewed distributions.
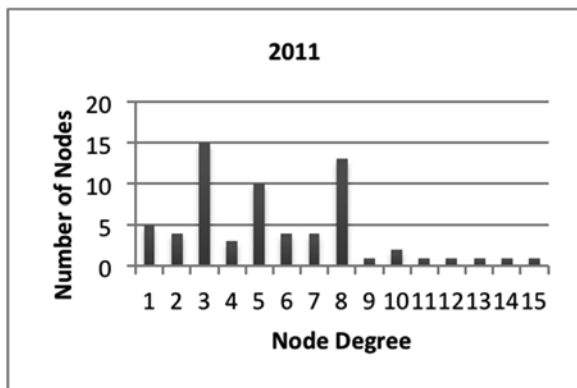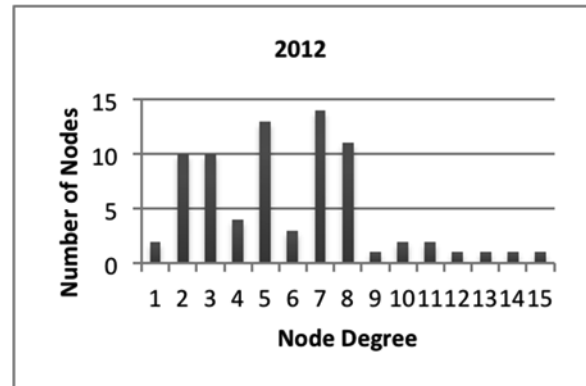

Figure 4: 2011 Degree Distribution

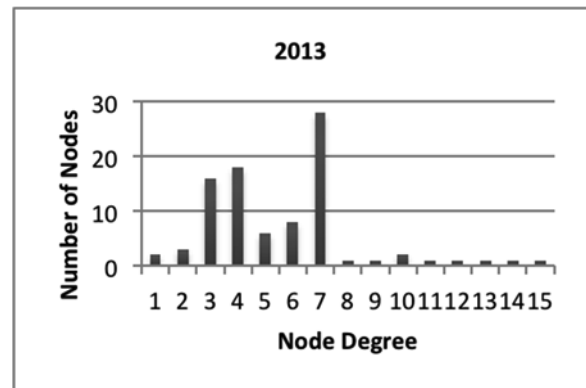**Figure 5: 2012 Degree Distribution**

Figure 6: 2013 Degree Distribution

Interestingly, the clustering coefficient drops significantly during the 2014 year to 0.541. The distribution of node degrees during 2014 (Figure 7) year appears normal, so this is a random graph. The distributions of node degrees for 2015-2018 do not follow either of the distributions that delineate a random or preferential attachment network. The 2015 and 2016 (Figures 8 and 9) networks have bimodal distributions and very high clustering coefficients at 0.859 and 0.899, respectively. Therefore, the 2015 and 2016 networks are small world networks.
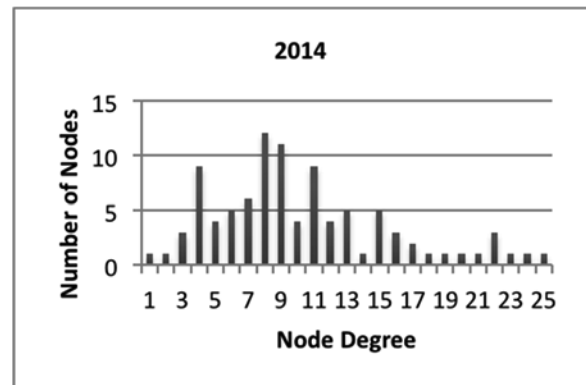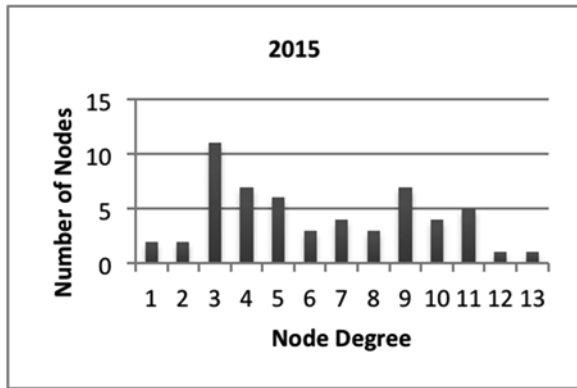

Figure 7: 2014 Degree Distribution
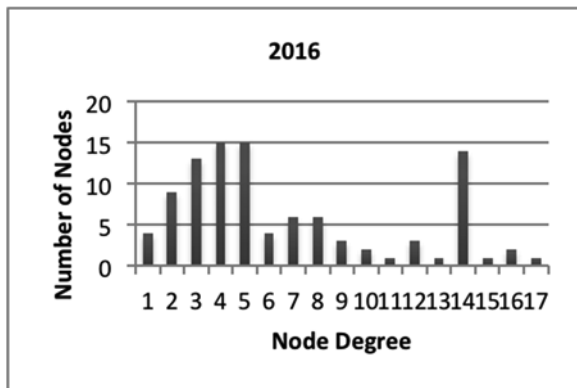
Figure 8: 2015 Degree Distribution

Figure 9: 2016 Degree Distribution

The 2017 (Figure 10) network topology is the most divergent of all the years as the distribution is bimodal. However, the clustering coefficient drops approximately 26 percent during this year, indicating that the nodes are less likely to be attached to the nodes with whom their project mates are working (i.e., cliquishness). The 2017 network represents a deviation in the norm, as the 2018 (Figure 11) network seems to follow the pattern established in 2015 and 2016 as it has both a bimodal distribution and a very high clustering coefficient (0.929).
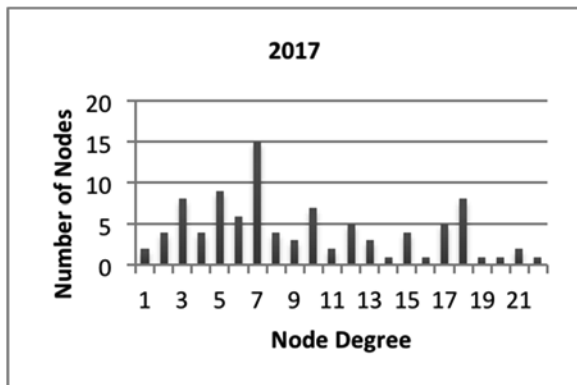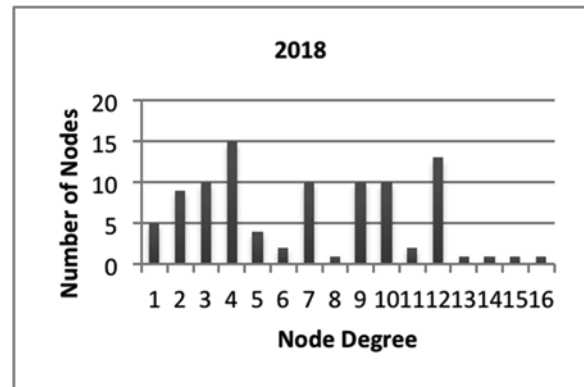

Figure 10: 2017 Degree Distribution

Figure 11: 2018 Degree Distribution

This analysis will focus on degree centrality for node level analysis. In 2010, the nodes with the highest degree centrality (in parentheses) were Nodes FC (22), MS (14), and AG (12). FC and MS were both permanent members and AG was a post-doc. In 2011, the nodes with the highest degree centrality were Nodes FC (34), MS (21), and AB (18). Here we see that FC and MS maintain their top positions as most central nodes, but AG dropped to the 34th spot in 2011. The top three individuals were all permanent members of the organization. In 2012,, the nodes with the highest degree centrality were FC (39), LK (25), and AG (23). Here, we see that LK rose to the second position in degree centrality. Despite its high degree centrality in 2011, AB fell two spots during 2012. Finally, we see that AG rose again to take the third highest spot. Like FC, LK was a permanent member. In 2013, the nodes with the highest degree centrality were FC (49), MS (24), and AG (23). During this year, LK fell to the fourth spot. During this year, the degree centrality for FC rose greatly, surpassing MS by 50 percent. In 2014 (Figure 7), the nodes with the highest degree centrality were FC (42), HC (41), and MS (30). During this year, FC's degree centrality fell slightly and HC (a permanent member) rose to the second spot with a degree centrality almost equal to FC, pushing MS to the third place. The 2014 network graph (see Figure 1) is visually more complex than the other networks, representing a potential climax in the organizational history.

The year 2015 represents a turning point in the organization. In 2015, there was a significant shift in the network, perhaps commensurate with the idea of complexity catastrophe [10 and 11] in which an organization becomes weakened by complexity and must readjust in order to survive. Between the years 2010-2014, there was consistent growth in the network, yet in 2015, there is a significant drop in membership. Many old members stepped back in their participation during this year and many new nodes

came forward. The nodes with the highest degree were VB (18), GA (14), and FC (12). Here we see three permanent members in the top three degree centrality, but the scores are much lower than in past years. This reflects a huge overhaul of the organization in which projects were re-evaluated and either consolidated with other projects or cut from the organization altogether. This year also saw a large decrease in participants over the course of one year. In 2014, there were 95 participants and the number fell to 56 in 2015. It is interesting that during this large overhaul, the director of the program (FC) fell to the third spot in degree centrality during this year. It is possible that FC took a step back from participating in as many individual projects in order to focus on administrative tasks with the re-organization. In 2016, the nodes with the highest degree centrality were FC (36), RS (22), and JW (22). During this year, many new projects were introduced and a new research center joined the organization. This was another year of shifting membership and project participation. During this year FC participated more in the projects and resumed his place as the most central member of the organization. In 2017, the nodes with the highest degree centrality were FC (47), WG (26), and JW (26). This year we see FC's degree centrality rise again to his pre-2015 score. This is the first appearance of WG in the top three most central nodes. WG is a permanent member with a high level role in his home organization. In 2018, the nodes with the highest degree centrality were FC (37), JW (23), and YR (22). FC and JW again represent some of the most central members but YR, also a permanent member, rose to the top three most central nodes for the first time.

5. Conclusions and Lessons Learned

The network evolved significantly through the years and changes to the organizational functioning definitely reflected in the network. The largest disparity at the node level occurred between 2014 and 2015. In the 2014 graph, there is only one component as the network is fully connected. The resulting network graph (see Figure 1) reflects an increase in network complexity consistent with the literature [8 and 10]. In 2015, there was a shift in membership. As a result of incorporating three new research centers, many new nodes were added and older nodes did not participate in projects at all. This reflects the reorganization of the projects during that year. After examining the number of products (such as papers and software development) for each year, initial results indicated that productivity generally dropped throughout the years.

Going into 2016, there was a huge influx of older nodes returning to projects and even more new nodes. Members of the newly added research centers became more involved in existing projects and created new ones. At the network level, the 2017 network represented a significant deviation from the prior two years. In some ways, the 2017 graph followed the trajectory set forth after the 2015 reorganization. The size of the network hovered just near 100 members and was becoming more connected between 2016 and 2017. That being said, 2017 was unique in that there was a 25 percent dip in clustering coefficient during this year. In 2018, the clustering coefficient rose again in accordance with prior years. One additional finding of this study is the high concentration of permanent members and relatively low concentration of graduate students and post-docs. One of the goals of this IVRO is to foster collaboration and provide education for graduate students and post-docs, but there are very few involved compared to the permanent members. Therefore, the network is revealing a potential problem in the organization that members either are not actively recruiting junior members to participate or the organization is not providing graduate students and post-docs the resources they need to be able to participate effectively.

One potential limitation of this project is that some teams place more importance on providing detailed information on progress for the annual report. Future research should explore possible reasons for this disparity and provide suggestions to the organization for how they can improve. In addition, future research should also examine how the network characteristics relate to the productivity of the teams since the annual report includes information about the products and achievements of the projects during that year. One might also examine the effect that multiple team membership has on the productivity and satisfaction of the members.

This study provides insights into the evolution of a network composed of purely mutual ties. The evolutionary stages of this IVRO had some similarities to Bryant and Monge's [10] community evolution stage model but there were a few differences that may point to key characteristics of evolutionary processes in this type of organization. One important difference of note is that the emergence and maintenance stages of evolution are somewhat difficult to distinguish using network density as an indicator. This is perhaps due, in part, to the fact that fewer organizations were involved and therefore, many collaborators were likely already familiar with one another and worked together prior to the emergence of the IVRO. It is possible that, due to this fact, the organization had a higher network density in the beginning than it would had

the organization been composed of more than two organizations. It is also possible that the absence of competitive ties in this network played a role in this finding. Future research should attempt to identify the cause of this finding and see if it reflects in the evolutionary processes in other similar organizations.

6. Acknowledgement

7. References

[1] R.B. Duque, M. Ynalvez, R. Sooryamoorthy, P. Mbatia, D.S. Dzorgbo, and W. Shrum, "Collaboration Paradox: Scientific Productivity, the Internet, and Problems of Research In Developing Areas", Social Studies of Science. 35(5), 2005, pp. 755-785.

[2] S. Lee and B. Bozeman, "The Impact of Research Collaboration on Scientific Productivity", Social Studies of Science, 35(5), 2005, pp. 673-702.

[3] D.H. Sonnenwald, "Scientific Collaboration", Annual Review of Information Science and Technology, 41(1), 643-680.

[4] A. Carusi and T. Reimer, Virtual Research Environment Collaborative Landscape Study, JISC, Bristol, pp. 4-106

[5] A. Abbasi, K.S. Chung, and L. Hossain. "Egocentric Analysis of Co-Authorship Network Structure, Position and Performance", Information Processing & Management, 48(4), 2012, pp. 671-679.

[6] E. Otte and R. Rousseau, "Social Network Analysis: A Powerful Strategy, Also for the Information Sciences", Journal of Information Science, 28(6), pp. 441-453.

[7] F.J. Acedo, C. Barroso, and J.L. Galán, "Co-Authorship in Management and Organizational Studies: An Empirical and Network Analysis", Journal of Management Studies, 43(5), 2006, pp. 957-983.

[8] N.J. Cooke and M.L. Hilton, Enhancing the Effectiveness of Team Science, National Academies Press, Washington D.C., 2010.

[9] A.H. Van de Ven, "On the Nature, Formation, and Maintenance of Relations Among Organizations", Academy of Management Review, 1(4), 1976, pp. 24-36.

[10] J.A. Bryant and P.R. Monge, "The Evolution of the Children's Television Community 1953-2003", International Journal of Communication, 2, 2008, pp. 160-192.

[11] B. McKelvey, "Avoiding Complexity Catastrophe in Coevolutionary Pockets: Strategies for Rugged Landscapes", Organization Science, 10(3), 1999, pp. 294-321.

[12] J. Scott, Social Network Analysis: A Handbook, SAGE Publications, London, 2000.

[13] D. Easley and J. Kleinberg, Networks, Crowds, and Markets: Reasoning About a Highly Connected World, Cambridge University Press, New York, 2010.

[14] K. Zweig, Network Analysis Literacy: A Practical Approach to the Analysis of Networks, Springer, Wien, Austria, 2016.

[15] A.L. Barabási, Linked: The New Science of Networks, Perseus Publishing, Cambridge, MA, 2002.