Many objective optimization and complex network analysis

Maulana, A.

Author: Maulana, A.
Title: Many objective optimization and complex network analysis
Issue Date: 2018-12-05

6

Towards Many-Objective Optimization of Eigenvector Centrality in Multiplex Networks

6.1 · Introduction

Identifying a set of key players in a network is an important research problem in many disciplines:

- In trading economy, it is important to know which countries are central in trade routes and networks of commodities and need to be stable in order to guarantee the long-term economic sustainability of economic networks.
- In political campaigning, it is important to identify a key player for reaching a large number of potentially interested people or people that should be made aware of a news item or political idea.
- In biology, ecosystems can be understood as networks of organisms. For instance, when maintaining a forest, it is critical to know which trees or organisms are most important to keep the forest in a healthy state by protecting them. There can be multiple networks that need to be considered, such as food webs, and signaling networks for communication and finding mating partners.

In the above problems, each node participates in multiple networks and high centrality in one network might not imply high centrality in another network. To better understand centrality concepts in such problems, in this work we will focus on a special class of networks – so-called multiplex networks – which are sets of networks which share the same set of nodes but differ in their links. As practical application domain We consider networks defining trade in different types of commodities.

Our paper presents the first step in the analysis of such multiplex networks by means of network centrality maximization, where network centrality is the most influential node in the network. We consider the optimization of network centrality in different layers (edge sets) as the objective functions. Optimizing several (3; 4) objectives simultaneously can be addressed by multi-objective optimization and many (> 3) objectives by many-objective optimization resulting in a high dimensional Pareto front. Since such high dimensional Pareto fronts are difficult to analyze, we also compute the Pareto fronts in pairwise different layers and analyze relationships between the objectives.

Layers can be in conflict with each other, meaning that they yield very different structures of centrality. They can be also complementary, meaning that maximizing the centrality in one layer also maximizes the centrality of the other layer. In this case, it is possible to merge the layers without losing essential information. Finally, it is also possible that the maximization of centrality of one layer does not affect the optimization of the centrality of another layer, in which case the problem could be easily decomposed.

6.2 • Related Work

Many-objective optimization is to optimize many objectives simultaneously; in this direction, various approaches have been developed. Some of them aim at reducing complexity, such as Objective Reduction in Many-objective Optimization: Linear and Nonlinear Algorithms [9], Reducing Complexity in Many-Objective Optimization Using Community Detection [39], and Objective Reduction Based on (Non Linear Correlation Information Entropy) [18]. Other approaches are based on Evolutionary Multi-objective optimization (EMO) extended to deal with many-objectives, cf. [32]. Related to multi-objective and many-objective optimization for network analysis such as Multi-Objective Optimization to Identify Key Players in Large Social Networks [25]. Some researchers did Multi-Objective Optimization for community detection / network clustering such as in [56], [8], [16], [31]. A Maximal Clique Based Multi-Objective Evolutionary Algorithm for Overlapping Community Detection [59], Overlapping Community Detection Through an Improved Multi-Objective Quantum-

Behaved Particle Swarm Optimization [36], and Community Detection From Signed Social Networks Using a Multi-Objective Evolutionary Algorithm [63].

In this chapter, we will follow the a similar approach to the multiplex network community detection (PaMoPlex) that was outlined in Chapter 5. Different to Chapter 5, not modularity, but centrality is considered as a maximization objective.

6.3 · Many-Objective Optimization of Network Centrality in Multiplex Networks

Our research approach in this chapter is to perform many-objective optimization of network centrality by computing and visualizing a matrix of Pareto fronts using two different approaches:

- By computing Pareto fronts for each pair of objectives and analyzing the results in a correlation matrix.
- By computing the Pareto front of the set of full-length objective vectors in $\mathbb{R}^m$.

Recall, that for every layer of the network one objective function is defined, which is to maximize the eigenvector centrality of that layer. In this way for a multiplex network G with layers $G_1, ..., G_m$ we define m as the number of objective functions.

Each node in the network is either dominated or non-dominated. A point of the node is said to be non-dominated if there is no other point which is better or equal in all criteria (all centralities in different layers) and better in at least one criterion (one layer). To compute the non-dominated subset from a finite set of n solutions, the algorithm by Kung, Luccio and Preparata is the fastest known approach [34], It accomplishes this task with a time complexity $O(n \log n)$ for $m = 2, 3$ and $O\left(n(\log n)^{m-2}\right)$ for $m > 3$.

We use the computation of pairwise Pareto fronts in order to understand and visualize the relationship of different layers with respect to centrality, i.e. whether or not and to which extent they share central nodes. Computing Pareto fronts of all objective functions can be used to easily recognize how many nodes are in the Pareto front. In order to compute a more fine-grained ranking of nodes, we compute the dominance rank of a node in the second analysis. Non-dominated solutions are of rank 1, solutions that are only dominated by rank 1 solutions are of rank 2, and so on. The computation of the rank only marginally increases the time complexity [13].

6.4 · Case Study and Implementation

6.4.1 · Analysis on Artificial Multiplex Networks

As an illustrative example on how to interpret results of the many-objective optimization based on network centrality, we started with computing the exact Pareto fronts for artificial multiplex networks. First, we did so for pairwise network layers, and then for all layers of the network.

The network was generated as a random graph based on the Erdős and Rényi model. Starting from a complete graph, each edge has a probability of $1 - p$ to be removed from the network. For a certain number of nodes ($m = 100$) and a certain probability ($p = 0.1$) we generated 11 layers for the experiment. We will denote them with g_1 to g_{11}. We chose eleven layers of the multiplex network, in order to later compare our results with the empirical economic trade multiplex network, which also has eleven layers.

The synthetic network based on Erdős-Rényi can be seen in Figure 6.1. From layer g_{11} of this artificial network, we compute the degree distribution. It is visualized as a histogram in Figure 6.2 and it has the shape of a binomial distribution, as it is typical for Erdős-Rényi graphs. Then we compute the centrality based on eigenvector centrality and the results are the vector of centralities of each node (100 nodes) in each layer. By dividing by the biggest eigenvalue (λ), for each layer, all values are normalized from 0 to 1.

For those 11 layers of the network, with eigenvector centralities of nodes in each layer, the next step is to compute non-dominated sets either by pairwise optimization or by many-objective optimization. The Pareto front is generated by full enumeration of node centrality of every node in every layer. Since the number of nodes in the Erdős-Rényi random graph is 100, we have 100 inputs for each layer. Each node has different centrality in different layers (objective functions) and by computing Pareto fronts for each pair of two layers, insight into the complementarity of layers can be gained.

For instance, in Figure 6.3, we can see that pairwise optimization for layers g_1 and g_2 yields six non-dominated nodes (marked by red points). The other nodes are not in the Pareto front but get a different color based on their dominance rank. Table 6.1 shows a partial list of node rankings of the Pareto front. There are several nodes in the second ranking, and the lowest ranking is 16. In Figure 6.4 all pairwise Pareto fronts and the correlation of the two objectives (layers) are compared with each other. The ability p = 0

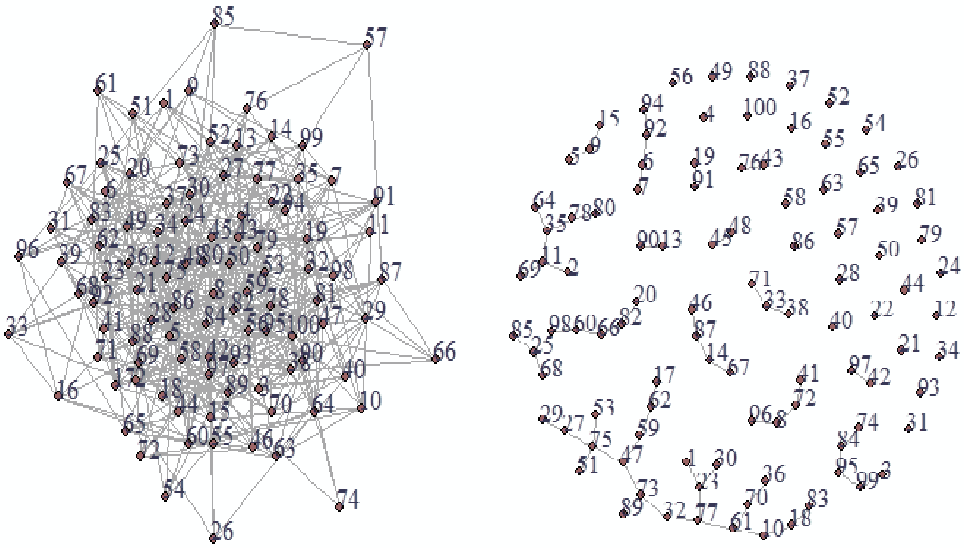

Figure 6.1 Erdős-Rényi random graph example with m = 100. The left network is generated with probability P = 0.1, and the one to the right with probability P = 0.01.

correlation strength of each pair is emphasized by color: Pink pairs have the highest correlation, green is in the middle, and yellow have the lowest correlation. There are no significant differences between the correlations, which can be attributed to the random nature of the network.

6.4.2 · Analysis on Trade Economic Multiplex Networks

The subsequent study shows that real multiplex networks exhibit a very different pattern as observed in random multiplex networks. An analysis for a real network from trade economic data is provided. We use the same data set that was used for multiplex modularity optimization in [41]. The data originates from network economy (trade data) using an import-export commodities network between 207 countries in 2011. (see [39]) The data represents the import-export relationships between some countries of the world, disaggregated for different traded commodities. This network can be defined as a multiplex network composed of many layers, where each layer is given by a different commodity group. The nodes are given by 207 countries. A link between two countries in the $i-th$ layer defined as weight exists if there is trade between

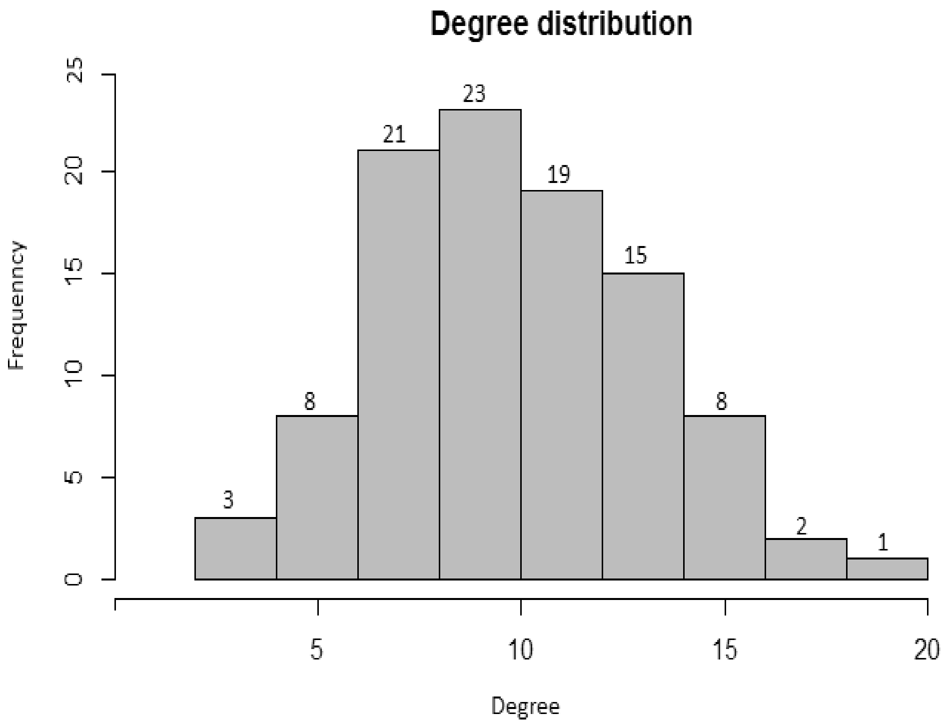

Figure 6.2 Degree distribution of Erdős-Rényi graph model for g_{11} (layer 11).

them in the $i-th$ layer, 6th commodity, for $i \in \{1, \ldots, 11\}$, denoting of 11 objective functions f_1 to f_{11}. Data are presented in matrix form: rows and columns represent countries, and the entries of the matrices are the volumes of trade. It is, therefore, a weighted multiplex network. In order to deal with weights, we use the matrix of weights instead of the adjacency matrix, where a weight of zero corresponds to the case the nodes are disconnected and the weights are proportional to the strength of connections. The general classification is based on 96 different commodities. The classification is performed by grouping together similar commodities; this procedure leads to 11 aggregated 'super-commodities'. The data represent the import-export relationships between countries of the world, disaggregated for different traded commodities. We have therefore a multi-layer (multiplex network) composed by many layers, where each layer is given by a different commodity. Each country represents a node of the layer, and a link between two countries in a given layer exists if there is trade between them in that commodity. The data used in the experiment is similar to the data used and

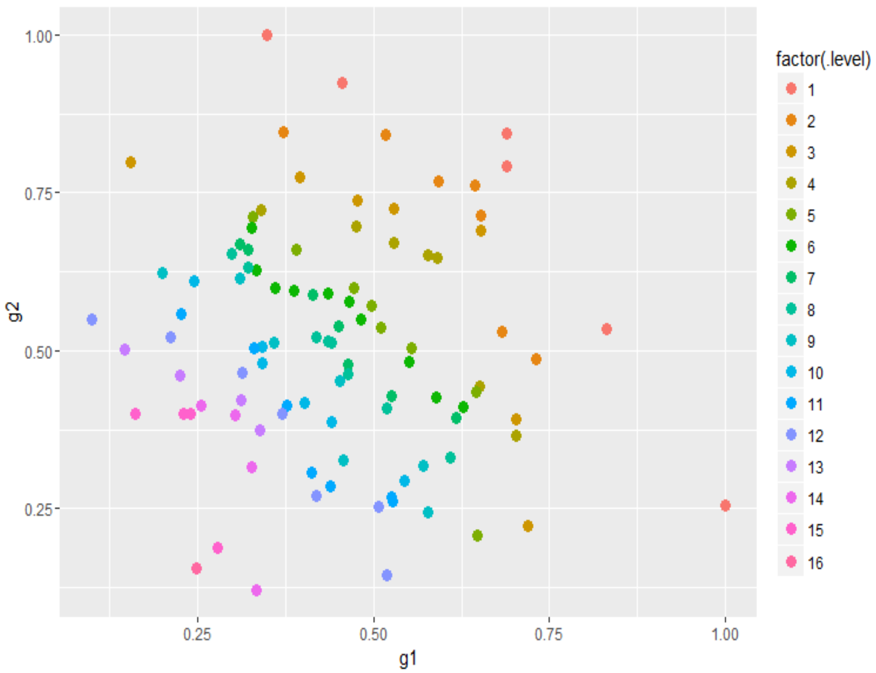

Figure 6.3 Pareto fronts of eigenvector centrality in network layer 1 and layer 2. Non-dominated solutions are shown by points with red colors in the first ranking.

described in the Chapter 5.

Similar to the random network with 11 layers of networks, here the trade multiplex network consist also of 11 layers. The first step is to compute the Pareto front of eigenvector centrality of each layer by full enumeration of 207 node centralities (for each country). As an example, Figure 6.5, shows the Pareto front between objective function f_7 (textiles) and f_{11} (optics and electronics).

Colors of the node centrality on the Pareto Front represent the ranking of the node in the Pareto Front, and the non-dominated solution in the Pareto front is represented by the first ranking from the set of solutions. To give a more clear insight regarding the ranking in the Pareto Front, we can analyze results from Table 6.2. For the non-dominated solution of pairwise optimization of f_7 and f_{11} we can see clearly that there are 5 countries that are non-dominated. They are China, Germany, Italy, England and USA.

From the Figure 6.6 we can see all country rankings in the set of solutions of pairwise optimization.

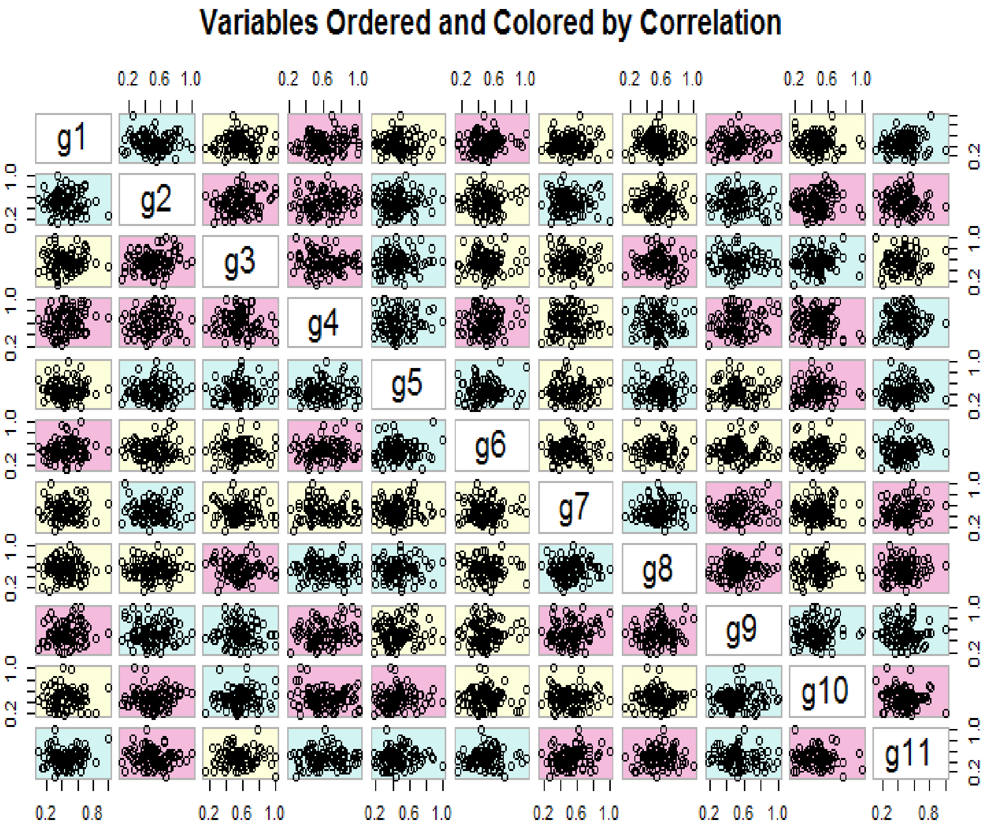

Figure 6.4 Scatter plot matrix for Pareto fronts of Eigenvector centrality with 11 layers of the Erdős-Rényi network model.

It becomes clear that the pairwise optimization for 11 network layers results in a different number of non-dominated solutions, each. There is at least one non-dominated solution and at most 5 non-dominated solutions for the pairwise optimizations. Single non-dominated solutions are USA in f_2 and f_3, France in f_1 and f_4, China in (f_5 and f_6), (f_5 and f_7), and (f_6 and f_7), and UK in f_{10} and f_{11}. There are mostly 2 non-dominated solutions such as China and France in f_4 and f_6, and also there are pairs of objectives with or 3, 4 and 5 non-dominated solutions. The maximal number of 5 non-dominated solutions only occurs in pairwise optimization in f_7 and f_{11}.

Finally, and most importantly, we compute by many-objective optimization the Pareto front for all objective functions (layers) considered together. We present results

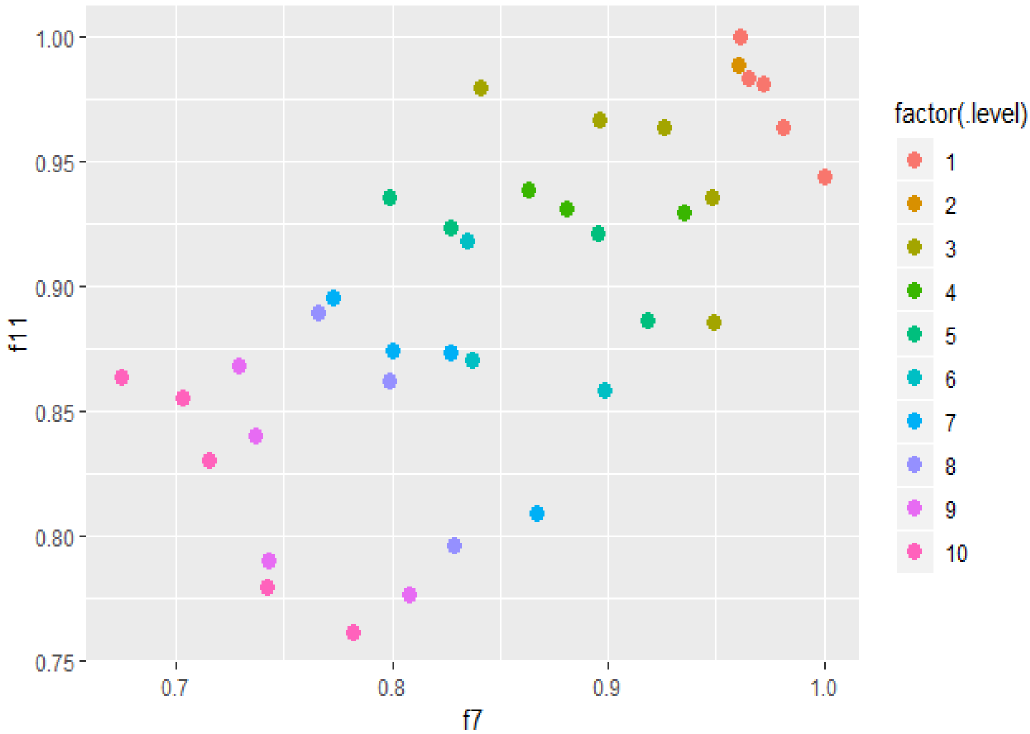

Figure 6.5 Pareto fronts ranking (big 10) of eigenvector centrality in network layer 7 and layer 11 of trade economic data.

in Figure 6.7.

In this figure, from the computation of the Pareto front for all layers together, we found there are 7 countries in the set of non-dominated solutions. They are China, France, Germany, Italy, India, UK, and the USA. Those countries are in the first ranking of the table. The second rank is shared by the countries Belgium-Luxembourg (BLX), Canada, Netherlands, Spain, Switzerland, and Turkey. In total there are 20 ranks and the last three ranks are only single countries Pitcairn Islands (PCN), West Sahara (ESH), Netherlands Antilles (ANT). The result shows that the centrality rank can serve as a rough indicator of the economic power of a country. Note, however, that a particularly strong centrality in one commodity type can be, in principle, the reason for the strong position in the ranking.

Comparing these results by means of pairwise optimization, we find that there are smaller numbers of non-dominated solutions, which is clear because solutions that are

$g1$	$g2$	$Node-label$	$.level$
0.45475356	0.9234941	6	1
0.6897551	0.7932208	19	1
1	0.254959091	43	1
0.6890175	0.8452013	49	1
0.8307032	0.5346712	59	1
0.3492441	1	70	1
0.644231	0.7625479	23	2
0.5920705	0.7692097	24	2
0.51719	0.8429117	13	2
.	.	.	.
.	.	.	.
.	.	.	.
0.3339607	0.1214376	69	14
0.2786245	0.18691	33	15
0.22968245	0.3995689	16	15
0.1624571	0.4001076	26	15
0.2401259	0.3993044	76	15
0.2487949	0.1558329	61	16

Table 6.1 Pareto fronts ranking of Eigenvector centrality in pairwise optimization in layer 1 and layer 2 from Erdős Rényi random graphs

$f7$	$f11$	$Country$	$Node-label$	$.level$
$1.00E+00$	$9.44E-01$	CHN	39	1
$9.65E-01$	$9.83E-01$	DEU	71	1
$9.81E-01$	$9.63E-01$	ITA	90	1
$9.61E-01$	$1.00E+00$	GBR	197	1
$9.72E-01$	$9.81E-01$	USA	199	1
$9.60E-01$	$9.88E-01$	FRA	65	2
$9.26E-01$	$9.63E-01$	BLX	16	3
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
$9.20E-03$	$3.54E-02$	PCN	146	57
$6.77E-03$	$1.51E-02$	ESH	175	58
$2.98E-18$	$2.16E-18$	ANT	125	59

Table 6.2 Pareto fronts ranking of Eigenvector centrality in network layer 7 and layer 11 from trade economic data.

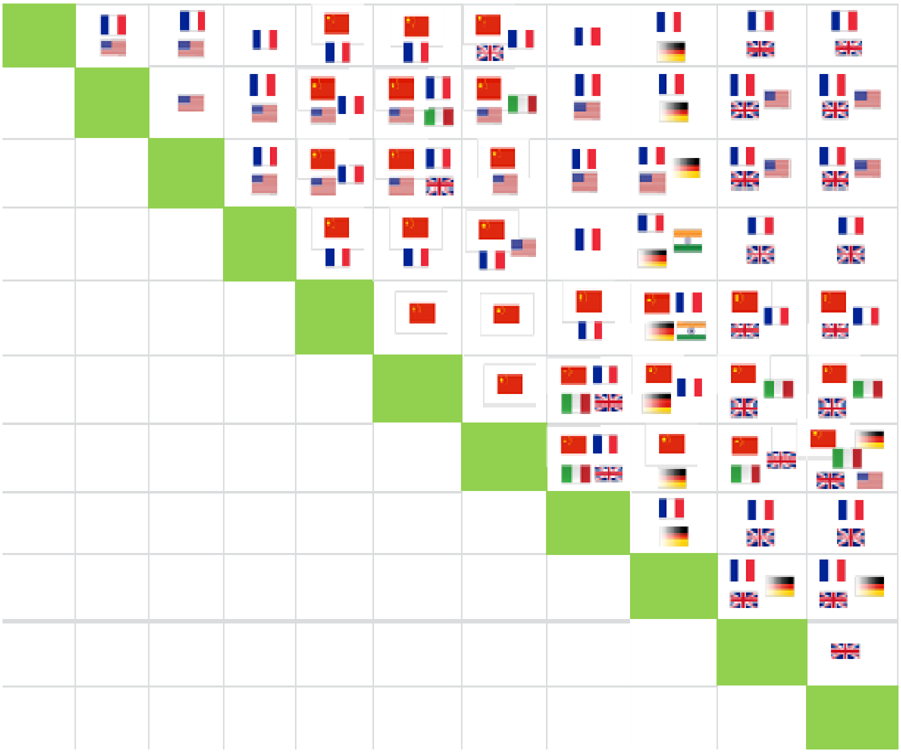

Figure 6.6 Scatter plot matrix for pairwise Pareto fronts of multiplex network trade economic data.Non-dominated solutions are represented by country flags as a vertex of the trade economic network.

non-dominated in only two objectives stay non-dominated when further objectives are added. There are also fewer rankings when all layers are considered. From Table 6.2, we obtain that there are 59 rankings for pairs of objectives (layer 7th and layer 11th), but there are only 20 rankings when optimize all layers together as it is shown in Figure 6.7.

6.5 • Summary

In this chapter, we discussed first results on the computation and analysis of Pareto fronts (set of non-dominated solutions) for eigenvector centrality in multiplex networks for the examples of Erdős Rényi random graphs and economic trade networks. As opposed to the maximization of modularity in previous work [41], the analysis of eigenvector centrality allows for using exact algorithms based on enumeration (all nodes of the networks) and efficient computation of non-dominated sets and dominance ranks of nodes. We discussed two analysis methods. They reveal different insight into the structure of the dominance relation and the relation between layers. The first method is to compute a Pareto front for every pair of layers. This shows whether similar nodes are central in the two selected layers or whether nodes are positioned very differently. Correlation analysis of the resulting Pareto front matrix can be used to quantify these results. In the analysis of the random graph no significant difference was observed, whereas, in the real-world networks, such as the trade network, the results differ from pair to pair and for some pairs single dominating countries could be identified. Secondly, the non-dominated solutions of the entire network can be computed as well as the dominance rank for all solutions. In the example of the trade multiplex network, the dominance rank is a rough indicator of how important a node is in the global trade network across different commodities. Analysis of the first ranks and last ranks of the networks yield plausible results with respect to this. The total number of non-dominated countries across all 11 commodity groups is, however, relatively small and consists of only 7 countries, all of them in the G20 countries (and 5 of them in the G8)[1]. For all other countries, there exist countries that are better or equal in all commodity centrality values and at least better in one centrality.

[1]India and China are non-dominated but not in the G8.

f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	Label	level
0.968	0.959	0.973	0.984	1.000	1.000	1.000	0.934	0.997	0.976	0.944	'CHN'	1
1.000	0.997	0.999	1.000	0.991	0.992	0.960	1.000	0.998	0.999	0.988	'FRA'	1
0.974	0.977	0.977	0.985	0.963	0.979	0.965	0.959	1.000	0.980	0.983	'DEU'	1
0.943	0.970	0.963	0.969	0.956	0.995	0.981	0.964	0.990	0.983	0.963	'ITA'	1
0.831	0.952	0.928	0.995	0.985	0.979	0.948	0.917	1.000	0.965	0.935	'IND'	1
0.932	0.959	0.973	0.977	0.956	0.994	0.961	0.980	0.995	1.000	1.000	'GBR'	1
0.989	1.000	1.000	0.994	0.977	0.986	0.972	0.998	0.993	0.973	0.981	'USA'	1
0.927	0.957	0.964	0.957	0.945	0.953	0.926	0.951	0.978	0.979	0.963	'BLX'	2
0.936	0.957	0.945	0.948	0.915	0.902	0.863	0.962	0.941	0.958	0.938	'CAN'	2
0.946	0.980	0.992	0.968	0.959	0.953	0.896	0.953	0.987	0.996	0.966	'NLD'	2
0.949	0.943	0.952	0.964	0.947	0.953	0.935	0.954	0.978	0.967	0.929	'ESP'	2
0.855	0.877	0.949	0.961	0.920	0.958	0.841	0.953	0.957	0.967	0.979	'CHE'	2
0.797	0.915	0.929	0.930	0.912	0.951	0.949	0.924	0.923	0.918	0.885	'TUR'	2
0.868	0.879	0.920	0.928	0.906	0.921	0.799	0.920	0.942	0.955	0.935	'AUS'	3
0.803	0.846	0.917	0.857	0.891	0.921	0.835	0.917	0.923	0.917	0.918	'AUT'	3
0.760	0.779	0.842	0.845	0.871	0.866	0.837	0.856	0.919	0.941	0.870	'CZE'	3
0.861	0.887	0.881	0.872	0.881	0.899	0.773	0.871	0.890	0.943	0.895	'DNK'	3
0.725	0.803	0.863	0.897	0.908	0.916	0.898	0.879	0.961	0.920	0.858	'IDN'	3
0.831	0.869	0.881	0.924	0.886	0.925	0.881	0.916	0.959	0.935	0.931	'JPN'	3
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.
0.029	0.028	0.051	0.055	0.056	0.054	0.039	0.044	0.053	0.084	0.064	'WLF'	17
0.024	0.019	0.008	0.018	0.034	0.017	0.009	0.016	0.048	0.070	0.035	'PCN'	18
0.006	0.000	0.000	0.005	0.005	0.000	0.007	0.012	0.016	0.036	0.015	'ESH'	19
0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	'ANT'	20

Figure 6.7 Pareto fronts of Eigenvector centrality 11 layers of the trade economic network. Non-dominated solutions are represented by ranking ("level") one in the the list.

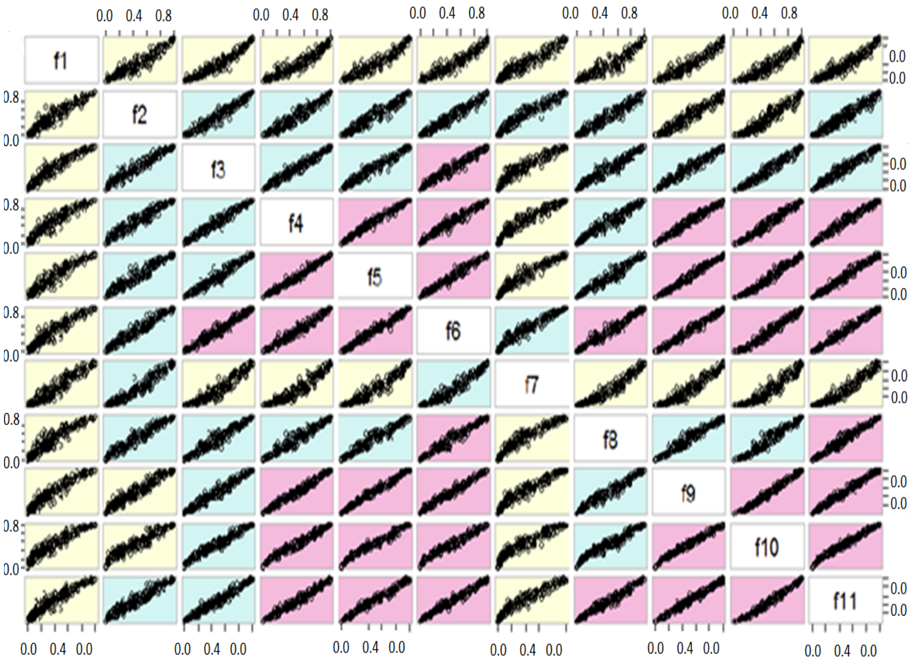

Figure 6.8 Scatter plot matrix of network correlations for the economic network.