Analysing Weighted Networks: An Approach via Maximum Flows

Markus Brede[1] and Fabio Boschetti[2]

[1] Centre for Complex Systems Science, CSIRO Marine and Athmospheric Research, GPO Box 284, Crace Canberra, ACT 2601, Australia
Markus.Brede@Csiro.au
[2] CSIRO Marine and Athmospheric Research, PB 5, Floreat, WA 6913

Abstract. We present an approach for analysing weighted networks based on maximum flows between nodes and generalize to weighted networks 'global' measures that are well-established for binary networks, such as pathlengths, component size or betweenness centrality. This leads to a generalization of the algorithm of Girvan and Newman for community identification. The application of the weighted network measures to two real-world example networks, the international trade network and the passenger flow network between EU member countries, demonstrates that further insights about the systems' architectures can be gained this way.

Keywords: Complex Networks, Weighted Networks.

1 Introduction

Describing systems of many interacting elements as networks has lead to considerable advances in the understanding of complex systems. Remarkably, such an analysis seems to indicate that systems from many different contexts, be it biology, ecology, sociology or even human-designed systems, share a plethora of common characteristics. Thus they can be roughly classified by the structure of their interaction topology. Common features are scale-free degree distributions, large degrees of clustering and modularity, small sizes and particular degree mixing patterns [1,2,3].

So far most of this analysis has treated networks as binary. However, in many applications such as transport networks [4,5] financial networks [6,7,8] collaboration networks [9], networks of metabolic fluxes [10] or gene regulatory and protein interaction networks [11,12] networks are weighted. Moreover, in some cases the distribution of link weights has been found to be very skewed, ranging over several orders of magnitude [4,7,10]. As an extreme example, imagine two pairs of cities, one coupled by a highway with a flow capacity of a thousand cars a day and the other one connected by a gravel road with a flow capacity of 10 cars a day. Should both connections be treated in the same way? Other studies indicate that strong and weak links have different roles in the systems' organization [13,14]. Naturally, this raises a number of questions: first, whether

J. Zhou (Ed.): Complex 2009, Part I, LNICST 4, pp. 1093–1104, 2009.

a binary representation of such systems is justified; second, whether more accurate information could be obtained by analyzing the weighted networks in its natural form; and third whether the general properties discussed above (degree distribution, clustering, etc.) are also found in weighted networks. In order to investigate these issues means of translating binary graph theoretical characteristics to weighted networks are needed.

The recent literature shows several efforts in this direction [4,5,11,15,16,19,20] which can broadly be grouped into three lines of research. In one approach, the 'Ensemble Approach' [5] link weights w_{ij} are mapped to probabilities $p_{ij} = F(w_{ij})$ for the presence of binary links between nodes and weighted network characteristics are defined as averages of binary quantities over the ensemble of networks defined by the connection probabilities p_{ij}. While this idea provides a nice general principle, there are two shortcomings. First, the definition of the map F is arbitrary; in [5] its definition requires the introduction of a lower cutoff probability whose setting is problematic due to the fact that weight distributions are often found to be very skewed with many very weak and few strong links. Small changes in this cutoff then determine the strength of the impact of all weak links —and hence the impact of a major fraction of the total link weights in the network— and can thus strongly influence the network measures. Second, network measures will typically have to be calculated by Monte Carlo simulations, such that the analysis of large networks can be very time-consuming and error-prone, if weak links are located in critical positions.

In the other line of research, differently motivated local weighted measures, mainly vertex strength, clustering coefficients or weighted nearest neighbour degrees, have been introduced [4,11,15,16]. From this approach there does not appear to be an obvious generalization of non-local measures, such as the size of connected components or distances.

Third, Ref. [17] introduces a mapping of weighted networks to multigraphs. On the basis that link weights represent capacities this allows for the generalization of global measures, notably the betweenness centrality in [17]. In this work we follow the spirit of [17] and elaborate global network measures based on the notion that weights in the network represent capacities or maximum flows. Focusing on 'global' network characteristics, we elaborate a number of measures, particularly distances, betweenness centralities and the definition of component sizes. Also, the introduction of a weighted betweenness centrality allows for a refinement of the cluster partitioning algorithm of Girvan and Newman [22].

2 Weighted Global Network Measures from Flow Principles

In this section we develop a set of weighted network characteristics and demonstrate their usefulness to understand the structure of two example networks, the trade flow network in 2000 and the passenger flow network between EU countries in 2004. For this, our guiding idea is the analysis of maximum flows along links. For both examples flows have a natural meaning in terms of the underlying system, i.e. bilateral trade flows for the trade network and passenger flows in the passenger network. Both are undirected networks, but the generalization of the network measures we propose to directed networks is straightforward.

Let us consider a weighted network of N nodes as given by a matrix $\{w_{ij}\}_{i,j=1}^{N}$, where an entry w_{ij} gives the weight of the link between the nodes i and j. Since we interpret link weights as capacities it is reasonable to assume $w_{ij} > 0 \forall i, j$. The binary representation of the matrix W is given by the adjacency matrix B, where $b_{ij} = H(w_{ij})$, $i, j = 1, ..., N$, where $H(x) = 1$ if $x > 0$ and $H(x) = 0$ otherwise. In the following, measures for standard binary networks refer to the matrix $\{b_{ij}\}$, whereas measures for weighted network refer to the matrix $\{w_{ij}\}$. In this terminology the number of links is given by $L = \sum_{i<j} b_{ij}$ and the (total) link weight by $S = \sum_{i<j} w_{ij}$. The average weight of a connection is obtained as $\overline{w} = S/L$.

Often a better understanding of the pecularities of a given network becomes possible by comparing it to a suitable null model. For the case of weighted networks, such a model is given by the ensemble of weight-randomized networks, i.e. networks with the same binary links and the same link weights, but an uncorrelated arrangement of the latter. In the following, results will be compared to averages over the ensemble of weight-randomized networks.

2.1 Component Size

In binary networks, cluster or component sizes give the sizes of maximum sets of connected nodes. Since in a densely connected weighted network all nodes can typically be reached from each other, a definition equivalent to the one for binary networks would not provide much information. However, even though all nodes can be reached from each other, the flow that can pass through the paths connecting them may be different. Thus, a sensible definition for the strength of the connection from a node i to a node j is the amount of flow $F_{\max}(i, j)$ that can simultaneously be passed along all links from i to j. Consequently, the average pairwise flow

$$F = \frac{1}{S} \sum_{i<j} F_{\max}(i, j) \tag{1}$$

provides a measure for the overall transport capacity between all nodes in the network, a measure roughly corresponding to component sizes in binary networks. Measuring F in units of the average link weight in (1) then allows for a better comparison between networks of different total link weight and generates a dimensionless measure. One should note that, despite (1) is not the exact equivalent of the cluster size for binary networks, the general concept of cluster size is respected, since (1) corresponds to the average number of independent path between nodes.

2.2 Pathlength

In binary networks pathlengths $l(i, j)$ are measured as the minimum number of edges that need to be traversed to establish a path between nodes

$$l(i,j) = \min_{P(i,j)} L(P(i,j)), \tag{2}$$

where $P(i,j) = (p_1 = i, p_2, ..., p_l = j)$ denotes a path from i to j and $L((p_1 = i, p_2, ..., p_l = j)) = l - 1$ its length. As in the previous section, such a quantity does not carry much information about the system structure in almost fully connected weighted systems such as the ITN or PFN, since one has $L = 1$ for almost every pair of nodes. A more sensible measure for pathlength must take into account the differences between strong and weak links.

From the flow perspective, we obtain a natural generalization of the discrete pathlength via the introduction of transport rates. To elaborate this concept, we need to introduce weighted paths as sequences of nodes and link weights traversed, i.e. $P_w(i,j)=((p_1 = i, p_2, ..., p_l = j), (v_{p_1,p_2}, ..., v_{p_{l-1},p_l}))$ with $v_{p_i,p_{i+1}} \leq w_{p_i,p_{i+1}}$ is a weighted path composed of all or parts of the edges w_{p_1,p_2}, w_{p_2,p_3}, ..., etc. Given such a weighted path $P_w(i,j)$ from i to j, we define its capacity t as $t(P_w(i,j)) = \min_{1 \leq k < l} v_{p_k,p_{k+1}}$ and, assuming that the transport along each link takes one unit of time, we define the transport rate as $r(P_w(i,j)) = t(P_w(i,j))/L(P_w(i,j))$, where $L(P_w(i,j)) = l - 1$ is the path's discrete length. A direct translation of the discrete pathlength for binary networks to weighted networks is thus obtained from the maximum transport rate $t_{\max}(i,j) = \max_{P_w(i,j)} r(P_w(i,j))$ via

$$l'_w(i,j) = \overline{w}/t_{\max}(i,j). \tag{3}$$

The definition (3) reduces to (2) for unweighted binary networks. However, a major caveat of (3) is that the computation of $t_{\max}$ requires the evaluation of all paths between two nodes, which is computationally demanding and impracticable for large networks.

A way around this difficulty is to consider the maximum average simultaneous transport rate between nodes, i.e. the average transport rates when the maximum possible amount of flow is carried between the nodes in the optimal way. To elaborate this concept, let us define that two weighted paths $P_w(i,j) = ((p_1 = i, p_2, ..., p_l = j), (x_{p_1,p_2}, ..., x_{p_{l-1}p_l}))$ and $Q_w(i,j) = ((q_1 = i, q_2, ..., q_l = j), (y_{q_1,q_2}, ..., y_{q_{l-1}q_l}))$ are independant, if any (allowed) flow through a part of one does not impede any flow through any part of the other. This is the case if $x_{mn} + y_{mn} \leq w_{mn}$ for all links m, n that occur in P_w and Q_w. A set of independant paths is a set of paths that are pairwise independant. Again, this definition is the natural equivalent of the definition of path independance in binary networks.

Then, the optimum maximum transport from i to j can be constructed in the following fashion:

1. The set of all considered paths $S(i,j)$ is empty.
2. Find the shortest (in the discrete sense) path $P_w(i,j)$ from i to j that is independent from all paths already in $S(i,j)$.
3. If no path fulfilling the requirements in 1. is found the algorithm terminates. Otherwise, add P to S and continue with 1.

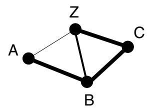

Fig. 1. This example illustrates the concept of weighted distances introduced in (5). Thin lines have weight 1, intermediate lines weight 2 and thick lines weight 3; thus the average link weight is $\overline{w} = 12/5$. Consider the distance between nodes A and Z; the maximum flow is $F_{\max}(A, Z) = 4$. The optimum maximum transport is obtained by transporting one unit directly via AZ, two units via ABZ and one unit via $ABCZ$. The corresponding independant paths are $P_1 = ((A, Z), (1))$, $P_2 = ((A, B, Z), (2, 2))$, and $P_3 = ((A, B, C, Z), (1, 1, 1))$, with the respective transport rates 1/1, 2/2 and 1/3, i.e. one has $r_{\text{avg.}}(A, Z) = 5/6$ and $l_w(A, Z) = 72/25$.

The sum $\sum_{P_w \in S(i,j)} t(P_w)$ gives the maximum simultaneous flow $F_{\max}(i, j)$ between i and j introduced in subsection 2.1. Thus,

$$r_{\text{avg}}(i, j) = 1/F_{\max}(i, j) \sum_{P_w \in S(i,j)} t(P_w) r(P_w) \tag{4}$$

is the weighted optimal average simultaneous transport rate between i and j. A measure for weighted pathlengths can be obtained from Eq. (4) by taking the inverse of the average transport rate measured in units of the average link weight $\overline{w}$

$$l_w(i, j) = \overline{w}/r_{\text{avg}}(i, j). \tag{5}$$

The concept is illustrated with an example in Fig. 1. An overall measure for distances between nodes in the weighted network is obtained by taking the average of (5) over all pairs of nodes.

Our definition of weighted pathlengths takes into account the discrete lengths of paths between nodes as well as their capacity. It is a measure of how much can be transported between nodes and how fast. However, similar to the definition of a component size in subsection 2.1, this also does not directly correspond to the unweighted measure for a pathlength in binary networks. Instead, definition (5) measures the average length of paths in the set of shortest independant paths between two nodes.

Eq. (4) can also be used as the basis for a definition of a weighted diameter. A sensible measure for the longest weighted path is $d_w = \max_{i \neq j} l_w(i, j)$.

To illustrate the concept we have measured the average weighted distances for the ITN and the PFN. In the ITN one finds $l_w = 1.47$ and in the PFN $l_w = 1.58$, compared to $l_w = 1.35$ and $l_w = 1.44$ for the null models, i.e. both networks are slightly larger than expected in a random link arrangement. For diameters one finds $d_w = 27920$ (between Andorra and Nassau) and $d_w = 45.3$ (Slovenia and Slovakia) for the ITN and PFN, respectively. Average diameters for the randomized networks are $d_w = 42400$ and $d_w = 12.3$, respectively. Thus, even though the average distance in the ITN is larger than expected, the diameter is considerably smaller than expected, i.e. for the ITN average pairwise

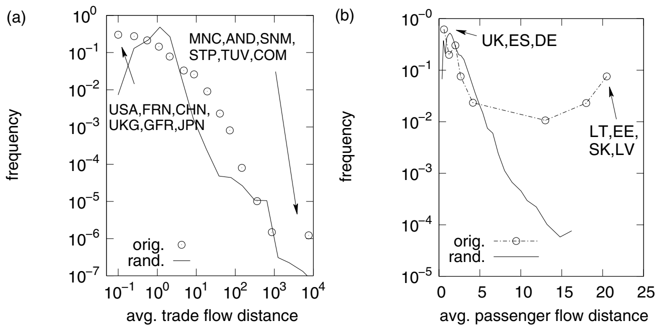

Fig. 2. Distribution of averages distances in the (a) ITN and (b) PFN. In the ITN the major industrialized countries and China have small average distances while small countries have large average distances. Andorra (AND) with $l_w = 11700$ has the largest average distance in the network. In the PFN the United Kingdom (UK), Germany (DE) and Spain (ES) have the smallest average distances while Listhuania (LT), Estonia (EE), Latvia (LV) and Slovakia (SK) have the largest.

distances are more 'concentrated' around the mean distance than in the respective randomized reference case. The opposite is found for the PFN, for which the distribution of distances is bimodal, see below.

Further information about both networks can be gained by analysing the distribution of average weighted pathlengths $\overline{l_w(i)} = 1/(N-1)\sum_j l_w(i,j)$ which are displayed in Figure 2. For both networks these distribitions are skewed, exhibiting pronounced differences between countries in the centre (short average weighted distances) and at the periphery (large average distances). For instance, the centre of the ITN is composed of a core comprising the US, Canada, Mexico, Germany, Japan, the UK, France and China while very small countries appear on the right tail. The distribution of average distances is more skewed than expected for random link arrangements, for which a peaked distribution of distances with most countries having similar average distances is expected, cf. Fig. 2a.

Even more in the PFN, a bimodal distribution of distances, clearly separating a 'core' from a 'periphery' is found (Fig. 2b). The UK, Germany and Spain form this clearly distinct core, while Slovakia and some Baltic countries are found at the periphery of the network. In contrast, in a random link arrangement a strongly peaked distribution of distances with only a few countries at the periphery would be expected.

2.3 Betweenness Centrality

In binary networks the betweenness centrality is a measure for the traffic through nodes (or links). It is usually defined as the number of shortest paths passing through a node (or link), i.e.

$$b(i) = \sum_{k,l} \delta(i,k,l) \tag{6}$$

with $\delta(i,k,l) = 1$ if a shortest path between the nodes k and l passes through the node i and $\delta(i,k,l) = 0$ otherwise. Again, as in the previous cases of discrete cluster size and pathlength, $b(i) \approx$ const. in almost fully connected weighted networks. That is, the discrete definition of betweenness centrality does not carry much information about densly connected weighted networks.

However, using the definition of a weighted pathlength in the previous subsection, a possible generalization of the discrete definition (6) to the weighted case becomes obvious. Let

$$b_w(i) = \sum_{k,l} \delta_w(i,k,l), \tag{7}$$

where $\delta_w(i,k,l)$ is the sum of the capacities of all paths through i used in constructing the optimum average simultaneous transport rate as in subsection 2.2. The definition (7) is thus an approximate measure for how much flow passes through nodes. The equivalent definition for a link centrality is obvious.

2.4 Vulnerability

One question about many distributed systems that is of considerable importance is that of vulnerability to random failure or targetted attack. The issue has been extensively studied for binary networks and recently also for weighted networks [18]. In Ref. [18] the vulnerability of the airtraffic network to random and targetted node removal is investigated. The authors investigate measures for structural damage to the network when nodes are removed according to different attack strategies. Different measures of structural system integrity are introduced, of which the most relevant one for the present study is the measure that computes the total link weight in the giant component. A major finding is that the topological structure of the binary network is a poor indicator of damage — the network can still be largely connected while node removals have already significantly impaired the system structure as captured by indicators that take account of the weighted architecture.

Though important, this result is not surprising: removing nodes based on a weighted centrality ranking will typically remove nodes of largest strength first, thus also removing larger amounts of link weight than when removing randomly chosen nodes. Thus naturally the amount of link weight remaining in the system decays faster than for random node removal. On the other hand, weak links tend to remain in the network and guarantee connectedness as long as many enough of them remain.

In the following, we analyse the integrity of the ITN subject to weighted link removal. This is equivalent to measuring system function when the overall trade volume is systematically reduced. For the analysis we use the measure of average pairwise maximum flows F of subsection 2.1 as an indicator of overall system integrity. Similar to [18] the fraction of nodes in the giant component N captures the system's topological integrity. More precisely, the ratios F_g/F_0 and N_g/N_0 are recorded. The indices indicate the measure when a fraction g of the total link weight has been removed.

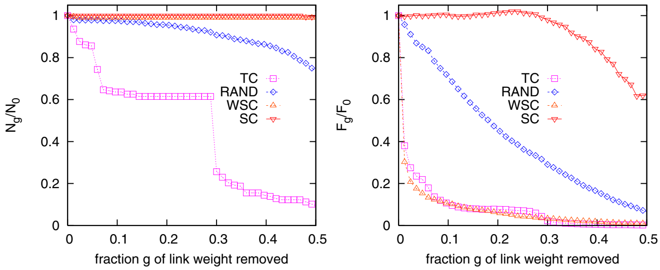

Fig. 3. Vulnerability of the ITN to random and targetted link removal. (a) fraction N_g/N_0 of the giant component remaining and (b) fraction of the maximum average pairwise flow (1) F_g/F remaining after the removal of a fraction g of the total link weight in the system. The labels indicate different removal strategies, TC removal based on ranking according to Eq. (7), RAND random link removal, WSC removal based on ranking according to the link centrality defined in [20] and SC removal according to standard unweighted centrality based ranking.

We consider the following strategies for gradual link removal, which, for practical reasons, is conducted in steps of $w_{\text{step}} = S/500$. For every step the removal procedure is repeated till an amount of w_{step} of total link weight has been removed. The ranking measures are recalculated after every removal step.

1. Random link removal (RAND).
2. Removal based on the centrality defined in Eq. (7) (TC).
3. Removal based on the weighted centrality defined in [20] (WSC).
4. Removal based on the standard link centrality (SC), that is purely based on the networks' topological structure.

Figures 3a and 3b compare simulation results for the topological and overall integrity of the system for various link removal strategies. One notes that for random and weight-based removal strategies the topological system integrity is largely unimpaired over a large parameter range whereas weighted indicators of system function already record major damage.

The data of figure 3b show a clear ranking of the attack strategies. Targetted removal according to an unweighted system indicator proves even less effective than random link removal. This observation underlines the importance of taking into account the arrangements of link weights to analyse the system. Most effective among the considered strategies are link removals based on the weighted centralities, where the weighted centrality measure introduced in [20] is slightly more effective for small amounts of removed weight, whereas a strategy based on the measure introduces in subsection 2.3 has larger impacts for large amounts of removed link weight. It is interesting to note that both strategies inflict comparable amounts of damage to the system in completely different ways. Note that for SC the topological integrity is completely unimpaired, wheras for TC considerable parts of the system are split off even when only relatively little link weight is removed, cf.Fig. 3a,b. Thus link removal according to SC impairs system function by thinning out one all-comprising giant component, whereas TC operates by gradually removing strategically placed links towards disconnecting the network.

Another observation appears particularly important: targetted removal of less than 3% of total trade volume can reduce the overall integrity of the system by around two thirds! Thus, even though the system is almost completely connected in a binary sense it can be classified as extremely fragile for targetted link removal.

2.5 Community Analysis and Modularity

The definition of centrality in the previous section can be used to generalize the algorithm of [22] to detect community structures in weighted networks. The algorithm of [22] dissects the network by sequentially removing the links with the largest centrality, thus gradually disconnecting the network. The basic idea behind the procedure is that more traffic flows along links connecting seperate communities, i.e. bridging links, than via those between nodes of the same community. In Ref. [20] a generalization of this procedure to weighted networks has already been suggested, thereby the traffic through a link (which is estimated from shortest paths on the binary network) is divided by the link weight. Then, links with the largest traffic per link weight ratios are removed. This procedure, however, does not account that weak links can become congested, consequently diverting a considerable fraction of the flow to strong links.

Our definition of centrality in (7) overcomes this problem. In more detail, to dissect weighted networks into communities we proceed as follows:

1. As long as the number of components is not equal to network size, calculate the weighted betweenness centrality of every link.
2. Remove the most 'overused' link, that is, the one with the largest weighted centrality over link weight ratio. Proceed with 1.

Figures 4 and 5 show the full dendrograms from the dissection of the ITN and the PFN into communities. The significance of a community division can

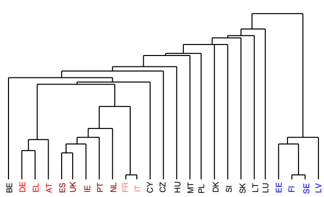

Fig. 4. Dendrogram from the dissection of the PFN into communities. Distances from the top represent the link weight removed to disconnect the network, i.e. highest branches represent early dissections and consequently most strongly disconnected communities. The most significant community division is the following split (1) 'central Europe': DE, EL, ES, FR, IT, AT, IE, NL, UK, PT (red), (2) Scandinavia and the Baltics: EE, FI, LV, SE (blue) and (3) the rest comprising 9 countries, each country as an isolated community (black). For this division the modularity is $Q = .57$.

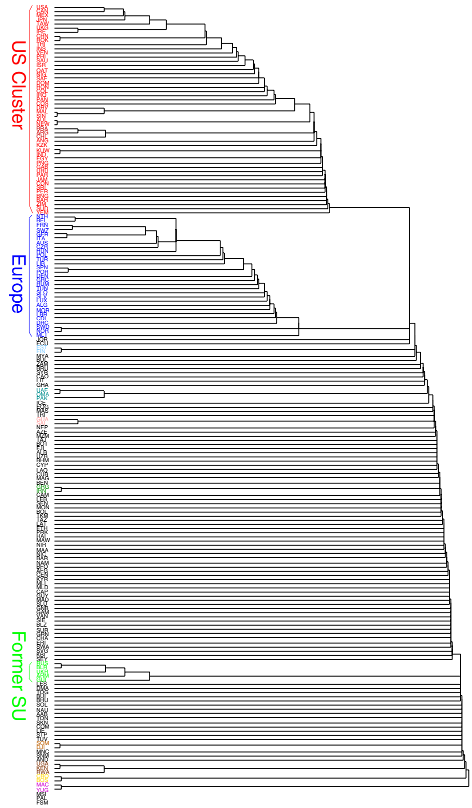

Fig. 5. Dendrogram from the dissection of the ITN into communities. The most significant community split distinguishes (1) a large clique of 50 mostly South and North American countries, but also some Asian countries and the UK, which is centred around the US, (2) a group of 29 (mostly) European countries, (3) a group of former Soviet Republics, (4) a multitude of very small groups (mostly with just one member) of other countries. The modularity for this split is $Q = .629$.

be classified according to its modularity Q [20], a measure for the ratio of the difference of link weight between members of the same groups and the expected fraction of link weight between group members in random arrangements and the total link weight in the network. The algorithm based on the traffic centrality of Eq. (7) yields $Q = .63$ whereas algorithms based on the standard centrality or weighted centrality of [20] only yield poorer community divisions with $Q = .182$ and $Q = .273$, respectively.

In the case of the PFN one notes that the most significant group division distinguishes three main sets of countries: (1) a 'core Europe' including 10 countries, (2) a block of four Scandinavian and Baltic countries, and (3) a set of nine countries, each defining a group of its own. Whereas the 10 countries of (1) form a clearly defined core containing almost all link weight, the set of countries summarized in (3) and to some extent the Scandinavian-Baltic cluster (2) form a clearly marked periphery. Nodes in this periphery have few connections between them and are also only loosely coupled to the network's core. The core itself can again be subdivided into three major groupings, one comprising Germany, Austria and Greece, a second containing Spain, the UK, Ireland, Portugal and The Netherlands and a third group made up by Italy and France.

For the ITN the most significant split distinguishes mainly two large trading blocks, one centred around the US and the other one comprising most European countries. Further, a clique of former Soviet Republics and several small cliques, mostly composed of only two or three countries are found. Both, the US centred and the European cliques are not homogeneous and can be further subdivided. For instance, the US clique contains a South American subgroup (comprising Brazil, Argentina and Chile), and several Asian subgroups (for instance China-South Korea or Malaysia, Signgapore and Vietnam). The bulk of Europe is centred around the Germany,France,Italy and Switzerland group, but also The Netherlands and Belgium and Sweden and Norway or Spain and Portugal form distinct subgroups.

3 Summary and Conclusions

By interpreting a network in terms of a transport system and introducing the concept of maximum flows between nodes, we proposed a natural extension of 'global' network measures traditionally used to study and characterise binary networks. Some, like vertex strength and clustering coefficient, can be seen as straightforward generalisations from binary to weighted networks. Others, such as cluster size, path length, component sizes and betweenness centrality, provide information which would not be available, or would be scarcely significant, if applied to the binary discretisation of a weighted network.

The measures we introduced allow to study a system accounting not only for the presence of links between constituents, but also for the strength of their interaction. We demonstrated the approach by analysing two real world networks: the International Trade Network and the Network of Passenger Flows between EU member countries in 2004. The two example networks have different size, and are characterized by link weights which span over several orders of magnitude with very skewed link weight distributions. The measures we discussed allowed to discriminate between these networks and reference 'null' models and to automatically detect several known features of the real systems.

References

1. Albert, R., Barabási, A.-L.: Rev. Mod. Phys. 74, 47 (2002)
2. Newman, M.E.J.: SIAM Rev. 45(2), 167 (2003)
3. Bocaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.-U.: Phys. Rep. 424, 175 (2006)
4. Barrat, A., Barthélemy, M., Pastor-Sarras, R., Vespignani, A.: Proc. Natl. Sci. U.S.A. 101, 3747 (2004)

5. Ahnert, S.E., Garlaschelli, D., Fink, T.M.A., Caldarelli, G.: Phys. Rev. E 76, 016101 (2007)
6. Onella, J.-P., Chakraborti, A., Kaski, K., Kertész, J., Kanto, A.: Phys. Rev. E 68, 056110 (2003)
7. Soromäki, K., Blech, M.L., Arnold, J., Glass, R.J., Beyeler, W.E.: Physica A 379, 317 (2007)
8. The data for the international trade network can be downloaded from, http://weber.ucsd.edu/~kgledits/exptradegdp.html We follow the procedure in [16] and reduce inconsistencies in the dataset by defining the link strength between countries as an average of reported imports/exports from the perspective of both countries
9. Newman, M.E.J.: Phys. Rev. E 64, 016131 (2001)
10. Almaas, A., et al.: Nature 427, 839x (2004)
11. Zhang, B., Horvath, S.: Stat. Appl. Gen. Mol. Bio. 4, Article 17 (2005)
12. Salwinski, L., et al.: Nucleic Acids Res. 32, D449 (2004)
13. Csermely, P.: Trends in Biochemical Sciences 29(7), 331 (2004)
14. Granovetter, M.: Am. J. Sociol. 73, 1360 (1973)
15. Onella, J.-P., Saramäki, J., Kertész, J., Kaski, K.: Phys. Rev. E 71, 065103R (2005)
16. Saramäki, J., Kivelä, M., Onella, J.-P., Kaski, K., Kertész, J.: e-print arXiv:cond-mat/0608670 (2006)
17. Newman, M.E.J.: Phys. Rev. E 70, 056131 (2004)
18. Dall' Asta, L., Barrat, A., Barthélmy, M., Vespignani, A.: J. Stat. Mech., P04006 (2006)
19. Holme, P., Park, S.M., Kim, B.J., Edling, C.R.: Physica A 373, 821 (2007)
20. Newman, M.E.J.: Phys. Rev. E 70, 056131 (2004)
21. Watts, D.J., Strogatz, S.H.: Nature 393, 440 (1998)
22. Girvan, M., Newman, M.E.J.: Proc. Natl. Sci. U.S.A. 99, 7821 (2002)
23. The data can be downloaded from, http://epp.eurostat.cec.eu.int
24. Newman, M.E.J.: Phys. Rev. Lett. 89, 208701 (2002)
25. Serrano, M.A., et al.: e-print ArXiv:cond-mat:0511035 (2005)
26. May, R.M., Levin, S.A., Sugihara, G.: Nature 451, 893 (2008)