# Widespread index

Vladimir Batagelj
Institute of Mathematics, Physics and Mechanics,
Department of Theoretical Computer Science,
Jadranska 19, 1 000 Ljubljana, Slovenia
and
University of Primorska, Andrej Marušič Institute,
Muzejski trg 2, Koper, Slovenia
e-mail: vladimir.batagelj@uni-lj.si

August 17, 2016 / 09 : 40

**Abstract**

Two indices for measuring the widespread of a node attribute value in a network are proposed and studied: the (simple) widespread index and the domination index. A computation of the proposed indices using the program Pajek is illustrated on two networks: the Class network and the US Airports network.

**Keywords:** social network analysis, widespread measure, dominating set.

## 1 Introduction

On January 12, 2016, Clement Levallois posted at SocNet the following message:

> I need to have a measure of how widespread is the distribution of a node attribute in a network. Let me explain:
>
> My nodes have a textual attribute, let's say "preferred flavor for ice cream". I would like to know to what extent the flavor "raspberry" is a value which is evenly distributed in the network, or to the contrary, just found in one community. A low value would mean that only nodes from a subregion of the network have this taste, a higher value would show an even distribution of the value across the whole network. I imagine that a difficulty is to account for the frequency of the attribute: if many nodes of the network have "raspberry" for the value of the attribute, it will tend to make this value distributed more widely.
>
> Any help or pointer on this would be very much appreciated!
>
> Thank you, Clement Levallois

In the paper we propose and study two indices for measuring the widespread of a node attribute value in a network. We also show how they can be computed in Pajek.

## 2 Simple widespread index

Let us start with some notions needed in the following discussion.

Let $V$ be a set of nodes and $S$ its subset with a given attribute value. Let further $\mathbf{N} = (V, L)$, $L$ is a set of links, be a simple network — no loops, two adjacent nodes are linked either by an arc (directed link) or an edge (undirected link). See the *Class* network in Figure 1. With $N(S)$ we denote the set of neighbors of the set $S$:

$$N(S) = \{u \in V : \exists v \in S : (v, u) \in L\}$$

and with $N[S] = S \cup N(S)$. We say that a node is *covered* by $S$ iff it is in the set $S$ or is a neighbor of some node from the set $S$. Therefore the set $N[S]$ is exactly the set of nodes covered from $S$. With $n = |V|$ we denote the number of nodes.

$D$ is a *dominating set* of a network $\mathbf{N} = (V, L)$ iff $N[D] = V$. $D$ is an *independent set* of a network $\mathbf{N} = (V, L)$ iff $D \cap N(D) = \emptyset$. For details see Haynes et al. (1998, p. 211).

Related to dominating sets is a *domination number* $\gamma(\mathbf{N})$

$$\gamma(\mathbf{N}) = \min\{|D| : D \text{ is a dominating set of } \mathbf{N}\}$$

for which it holds $n \geq \gamma(\mathbf{N}) \geq \lceil \frac{n}{1+\Delta} \rceil \geq 1$, where $\Delta$ is the largest outdegree in $\mathbf{N}$.

**Note 1.** Even better lower bound is $\gamma(\mathbf{N}) \geq k$, where $k$ is the smallest number such that

$$\sum_{i=1}^{k} (1 + d_i) \geq n$$

where $(d_i)_i$ is a sequence of outdegrees ordered in decreasing order.

Let $S$ be the set of nodes having a selected attribute's value. A simple measure of its widespread is the size of the set $N[S]$. For $S \neq \emptyset$ it holds $0 < |N[S]| \leq n$. Normalizing it we get a *simple widespread index*

$$W(S) = \frac{|N[S]|}{n}.$$

It is easy to verify that

**Proposition 1.** *The simple widespread index $W(S)$ has the following properties:*

*(a)* $0 \leq W(S) \leq 1$.

*(b)* $W(V) = 1$.

*(c)* $W(S) = 1$ *iff* $S$ *is a dominating set.*

*(d) if* $S_1 \subset S_2$ *then* $N[S_1] \subseteq N[S_2]$

*(e)* $W(S_1) < W(S_2)$ *iff* $|N[S_1]| < |N[S_2]|$.

# 3 Domination index

A problem with the definition of the simple widespread index $W(S)$ is that it doesn't consider the size of the set $S$. The set $N[S] \setminus S$ is the "expansion" of the set $S$ – a set of new nodes linked from $S$. A well widespread set $S$ has a large expansion. Let $D^*$ be a *minimum dominating set*, $|D^*| = \gamma(\mathbf{N})$. Then an alternative widespread measure of $S$ could be a normalized size of its expansion which we call a *domination index*

$$W^*(S) = \frac{|N[S] \setminus S|}{|V \setminus D^*|} = \frac{|N(S) \setminus S|}{n - \gamma}.$$

If $S = \emptyset$ we set $W^*(S) = 0$. Note that ($S^c = V \setminus S$)

$$N[S] \setminus S = (N(S) \cup S) \cap S^c = N(S) \cap S^c \ \cup \ S \cap S^c = N(S) \setminus S$$

It is easy to see that

**Proposition 2.** *The domination index $W^*(S)$ has the following properties:*

*(a)* $0 \leq W^*(S) \leq 1$ .

*(b)* $W^*(V) = 0$.

*(c) if $S$ is a minimum dominating set then* $W^*(S) = 1$.

*(d) if* $|N[S_1]| = |N[S_2]|$ *and* $|S_1| < |S_2|$ *then* $W^*(S_1) > W^*(S_2)$.

Most proofs are simple. Let us show that $W^*(S) \leq 1$. In general we have $N[S] \subseteq V$. Then there exists $P \subseteq V$ such that $N[S] \cap P = \emptyset$ and $N_+(S \cup P) = V$. Note that also $S \cap P = \emptyset$. If $N[S] = V$ we set $P = \emptyset$. Because $S \subseteq N[S]$ it holds

$$|N[S] \setminus S| = |N[S]| - |S|$$

From

$$n = |V| = |N[S \cup P]| \geq |N[S]| + |P|$$

we get

$$|N[S]| - |S| \leq n - |S| - |P| = n - |S \cup P|$$

Because $S \cup P$ is a dominant set $|S \cup P| \geq \gamma$ and therefore

$$|N[S]| - |S| \leq n - \gamma$$

or finally $W^*(S) \leq 1$.

The value $W^*(S) = 1$ can be attained also in cases not covered by Proposition 2.c . For example in the 4-cycle $\mathbf{C}_4 = (V, E)$, where $V = \{a, b, c, d\}$ and $E = \{(a : b), (b : c), (c : d), (d : a)\}$, we have $n = 4$, $\gamma = 2$, and for $S = \{a\}$

$$W^*(\{a\}) = \frac{|\{a, b, d\} \setminus \{a\}|}{4 - 2} = 1$$

These cases satisfy the following scheme.

Let $D^*$ be a minimum dominating set in a network $\mathbf{N}$ and $D_0 \subset D^*$ its subset such that the subnetwork induced by the set $V \setminus N[D_0]$ is a null graph – it has no link. Then for each $D$, $D_0 \subseteq D \subseteq D^*$ it holds $W^*(D) = 1$. Note also that each such $D$ can be extended to a minimum dominating set $D \cup (V \setminus N[D])$.

Unfortunately the problem of determining the domination number $\gamma(\mathbf{N})$ is NP-complete – there is no efficient algorithm to compute $W^*$ (Garey and Johnson, 1979, p. 190, problem GT2).

**Note 2.** Nodes with indegree 0 belong to every dominating set of a given network.

**Note 3.** Let $D_0$ be the set of all nodes with zero indegree. Then we get a better lower bound for $\gamma$ — it is $|D_0| + k'$, where $k'$ is the smallest number such that

$$\sum_{i=1}^{k'}(1 + d_i) \geq n - |N_+(D_0)|.$$

Since $V \setminus D_0$ can have zero degree nodes again, we iterate the process. The final $d_i$s are computed in the final $V \setminus D_0$.

**Note 4.** In some real life networks we have many "leaves" – nodes with indegree 1 and outdegree at most 1. For such nodes there always exists a minimal dominant set that contains their twin node – "roots".

To get an efficiently computable index we could replace $\gamma$ with 1 (it always holds $1 \leq \gamma$); or, even better, with some other lower bound $k$ (also $k \leq \gamma$). The *domination $k$-index* is defined as

$$W_k(S) = \frac{|N(S) \setminus S|}{n - k}.$$

**Proposition 3.** *For the domination $k$-index $W_k(S)$ it holds:*

*(a)* $W^*(S) = W_\gamma(S)$ .

*(b)* $k_1 < k_2 \Rightarrow W_{k_1}(S) < W_{k_2}(S)$.

*(c)* $k < \gamma \Rightarrow 0 \leq W_k(S) < 1$.

*(d)* $W^*(S_1) > W^*(S_2)$ *iff* $W_k(S_1) > W_k(S_2)$

*(e) if* $|N[(S_1]| = |N[S_2]|$ *and* $|S_1| < |S_2|$ *then* $W_k(S_1) > W_k(S_2)$.

# 4 Examples

## 4.1 The Class network

The *Class* network (see Figure 1) describes the borrowing of study materials among 15 social informatics students at Faculty of social sciences in Ljubljana (Pajek data sets, 2016). Besides the borrowing relation we have also information about the sex (man, woman) of each student. There are $n_m = 6$ men and $n_w = 9$ women. Let us determine the widespread of each sex group, Men and Women, in the network.

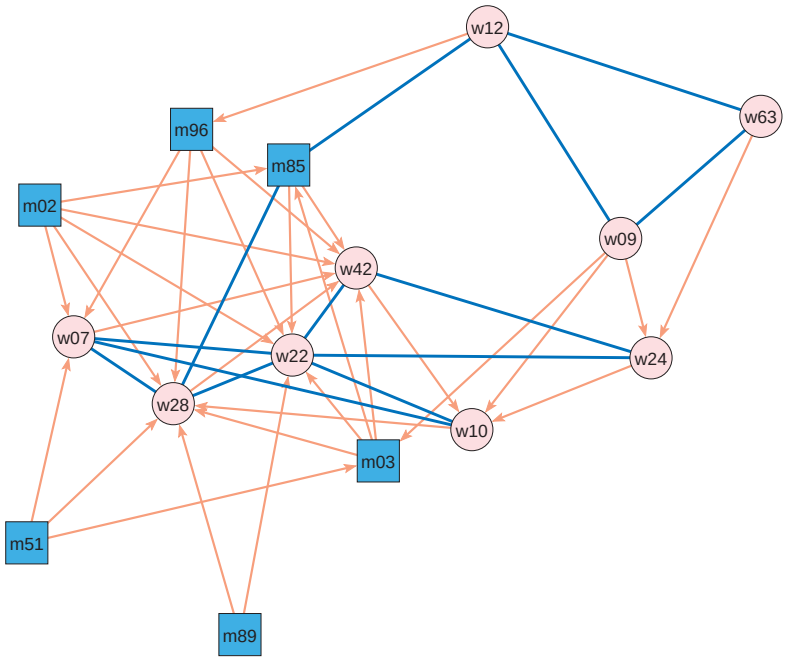


Figure 1: Class network

We first determine the numbers $k$ and $\gamma$. The outdegree of nodes m02, w22 and w09 is 5. Therefore by Note 1

$$(1 + \text{outdeg}(\text{m02})) + (1 + \text{outdeg}(\text{m02})) + (1 + \text{outdeg}(\text{m02})) = 18 > 15 = n$$

we get $k = 3$.

Applying Note 2 we see that $\{\text{m02}, \text{m51}, \text{m89}\}$ are in any dominating set. To cover the remaining uncovered nodes we need to add at least two additional nodes – for example, w12 and w24. We get $\gamma = 5$.

Using Pajek it is relatively easy (see Appendix A) to determine the sets $N[S]$ and $N(S) \setminus S$ and their sizes – needed for computing the indices. For the Class network we get:

| $S$ | Men | Women |
|---|---|---|
| $\lvert S\rvert$ | 6 | 9 |
| $\lvert N[S]\rvert$ | 11 | 12 |
| $\lvert N(S)\setminus S\rvert$ | 5 | 3 |
| $W(S)$ | $\frac{11}{15}=0.73333$ | $\frac{12}{15}=0.8$ |
| $W_3(S)$ | $\frac{5}{12}=0.41667$ | $\frac{3}{12}=0.25$ |
| $W^*(S)$ | $\frac{5}{10}=0.5$ | $\frac{3}{10}=0.3$ |

We see that $W(\text{Men}) = 0.73 < 0.8 = W(\text{Women})$ — women are more widespread than men. But, the set Men is smaller than the set Women and has larger expansion — men are more dominant in a network than women: $W_3(\text{Men}) = 0.42 > 0.25 = W_3(\text{Women})$ and $W^*(\text{Men}) = 0.5 > 0.3 = W^*(\text{Women})$ .

## 4.2 US Airports

As a non-toy example, let us consider the *US Airports* network (**?**). It consists of 332 airports and 2126 edges among them. There is an edge linking a pair of airports iff in the year 1997 there was a flight company providing flights between those two airports. Because the edge weights are not relevant for our example we set all weights to 1.

For this network it turns out that using Note 4 we can relatively easy determine its domination number $\gamma$. We call Outsiders the set $V \setminus N[\text{Roots}]$. Pajek commands for determining the set $N[\text{Outsiders}]$ are given in Appendix B. To determine a minimal dominating set $D^*$ we have to add to the set Roots a minimal set $D_o$ that covers the set Outsiders – light (green or yellow) nodes in Figure 2. Dark (red) nodes are neighbors of Outsiders that are covered by Roots.To cover an outsider we can use also some of its not outsider neighbor.

The solution is not unique. For example to cover the outsider 32 we can select any of the nodes 32, 31, and 33. To cover nodes 87, 66, 116 and 64 with a single node we have to select the node 65. Here is a minimal set

$$D_o = \{32, 55, 65, 97, 98, 141, 171, 262, 302, 320\}$$

The set of Roots contains 26 nodes + additional 10 nodes from $D_o$ give a minimal domination set $D^*$ with $\gamma = 36$ nodes.

Let us compute both indices for the set Leaves and the set 5_airports which consists of 5 nodes

$$5\_\text{airports} = \{8, 118, 248, 255, 261\}$$

that correspond to airports: Anchorage Intl, Chicago O'hare Intl, Los Angeles Intl, The William B Hartsfield Atlanta, and Dallas/Fort Worth Intl.

For the US Airports network we have $n = 332$ and $\gamma = 36$. We get the following table:

| $S$ | Leaves | 5_airports |
|---|---|---|
| $\lvert S\rvert$ | 55 | 5 |
| $\lvert N(S)\rvert$ | 26 | 218 |
| $\lvert N[S]\rvert$ | 81 | 218 |
| $\lvert N(S)\setminus S\rvert$ | 26 | 213 |
| $W(S)$ | $\frac{81}{332}=0.243976$ | $\frac{218}{332}=0.656627$ |
| $W^*(S)$ | $\frac{26}{296}=0.087838$ | $\frac{213}{296}=0.719595$ |

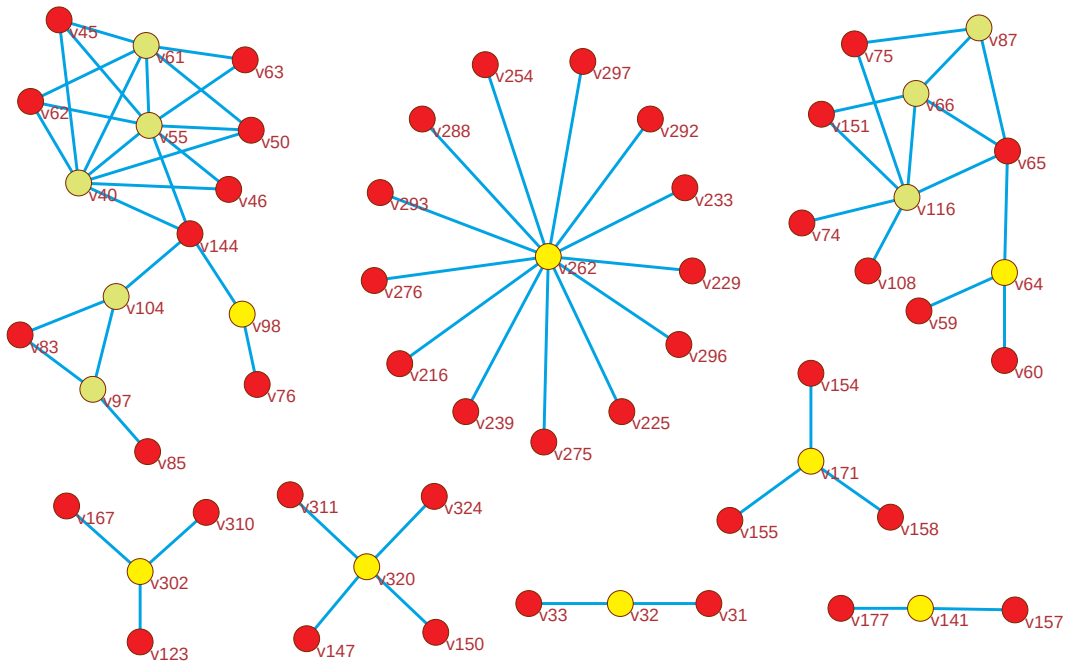


Figure 2: $N[\text{Outsiders}]$

For both indices their value for the set Leaves is smaller than the value for the set 5_airports. There are 55 leaves that cover 81 nodes; their extension contains onply 26 nodes. The selected 5 airports cover 218 nodes and have the extension with 213 nodes.

In Figure 3 a cut-out from US Airports network picture is presented. The light (yellow) node represents the Los Angeles Intl airport. The dark (red) nodes are airports linked with at least one of the selected 5 airports; and the gray (green) nodes are airports not linked to any of selected 5 airports.

# 5 Conclusions

In the paper we proposed two indices to measure a widespread of a subset of nodes in a given network. The simple widespread index measures a proportion of the set of nodes covered by a given subset. The dominance index considers also the "efficiency" of a given subset. The problem with the dominance index is that (it is believed that) no efficient algorithm exists for computing the network's dominance number $\gamma$. For comparing subsets inside the same network we can apply Proposition 3.d ($W^*$ and $W_k$ are determining the same ordering of subsets) and instead of $W^*$ use $W_k$ for some lower bound $k$.

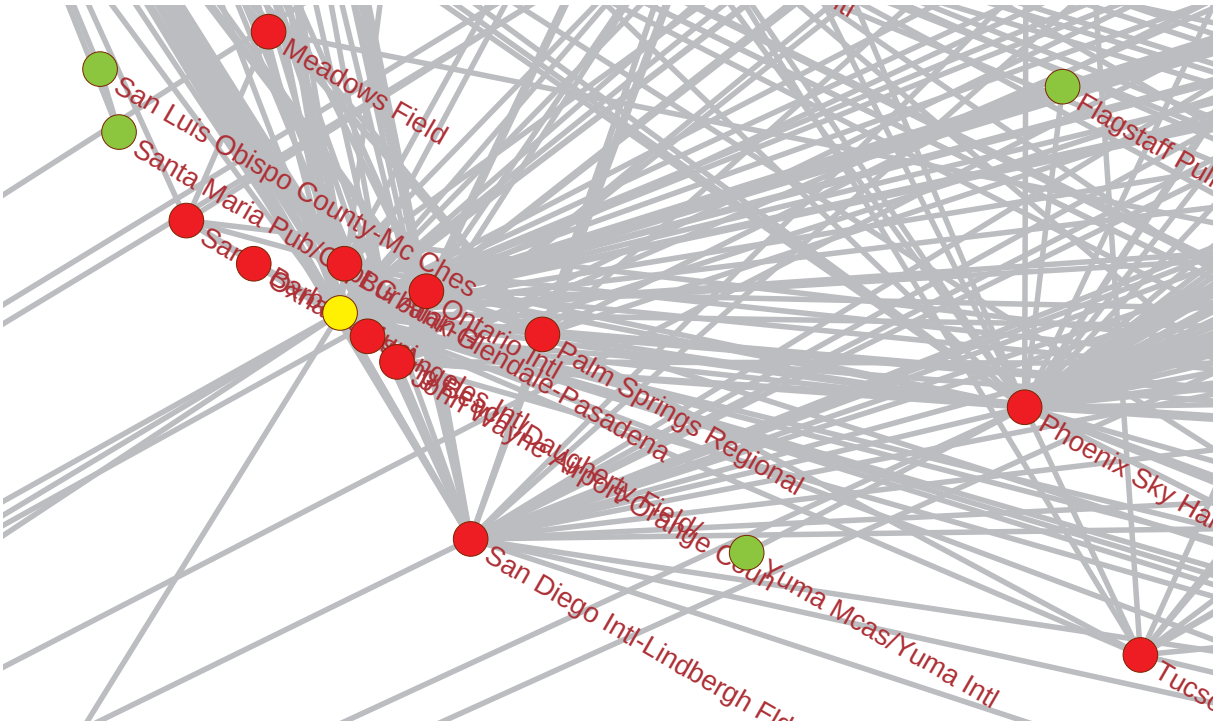


Figure 3: US Airports 1997 / 5_airports, a cut-out


## Acknowledgments

This work was supported in part by the Slovenian Research Agency (research programs P1-0294 and research projects J5-5537 and J1-5433).

The paper is an elaborated and detailed version of the talk presented at the Second European Conference on Social Networks, Paris, June 14-17, 2016.


# References


Batagelj, V., Mrvar, A. (2016), Pajek data sets:
Class network: http://vlado.fmf.uni-lj.si/pub/networks/data/test/class.net ,
US Airports network: http://vlado.fmf.uni-lj.si/pub/networks/data/mix/USAir97.net .

Batagelj, V. (2016), Widespread. https://github.com/bavla/widespread/ .

De Nooy, W., Mrvar, A., Batagelj, V. (2011), Exploratory Social Network Analysis with Pajek; Revised and Expanded Second Edition. Structural Analysis in the Social Sciences, Cambridge University Press.

Garey, M.R., Johnson, D.S. (1979), Computers and Intractability: A Guide to the Theory of NP-Completeness, W. H. Freeman.

Haynes, T.W., Hedetniemi, S., Slater, P. (1998), Fundamentals of Domination in Graphs. Marcel Dekker.

# A Computing widespread indices in Pajek

Here is a short description of a procedure to compute in Pajek the simple widespread index and the domination $k$-index. Setting $k = \gamma$ we get the domination index.

We first read the network into Pajek:

```
File/Network/Read [class.net]
```

In some way we determine the network's domination number $\gamma$ or its lower bound $k$. For the class network we get $\gamma = 5$ and $k = 3$.

Let $V = \{v_1, v_2, v_3, ..., v_n\}$. We assign to its subset $S \subseteq V$ the corresponding characteristic vector $\chi(S) = [h_1, h_2, h_3, ..., h_n]$ where $h_i = 1$ if $v_i \in S$, and $h_i = 0$ otherwise.

We partition the node set to sets Men and Women according to the node shape (square - man; circle - woman).

```
Network/Create Partition/Vertex Shapes
Partition/Binarize Partition [1]       % 1=man, 2=woman -> 0=woman, 1=man
```

Clicking on the Info button for partition we learn that the group 1 contains 6 men, and the group 0 contains 9 women.

Assume that we have active in Pajek registers: the network, the partition $S$ and the scalar $k$. Then both indices are computed as follows:

```
Network/Create New Network/Transform/Transpose 1-Mode [yes]
Partition/Copy to Vector
Operations/Network + Vector/Network*Vector [1,OK]
Vector/Make Partition/by Intervals/Selected Tresholds [0.5]
Vector/Create Scalar/Number                          % n
Partition/Binarize Partition [2]                     % N(S)
select partition S  as Second
Partitions/Max(First,Second)
Partition/Copy to Vector
Vector/Create Scalar/Sum
select scalar n as Second
Vectors/Divide (First/Second)
File/Vector/Change Label [W]
select partition S as First
Partition/Binarize Partition [0]
select partition N(S) as Second
Partitions/Min(First,Second)
Partition/Copy to Vector
Vector/Create Scalar/Sum                             % |N(S)-S|
select scalar n as First
select scalar k as Second
Vectors/Subtract (First-Second)                      % n-k
select n-k as Second
select |N(S)-S| as First
Vectors/Divide (First/Second)
File/Vector/Change Label [Wk]
```

This sequence of commands is saved as the macro `widespread`. It expects in active registers as "inputs" a network, a subset $S$ given as a binary partition (characteristic vector), and a scalar $k$. It returns both indices $W$ and $W_k$ (Batagelj, 2016). The macro `widespread` removes the auxiliary data. If you are interested in the intermediary results apply the macro `wide` instead.

# B Creating in Pajek a picture from Figure 2

We first read the US Airports network from the file `USair97.net` into Pajek. We follow a construction implied by Note 4 and produce the sets Leaves, Roots, Outsiders and $N_+(\mathrm{Outsiders})$:

```
Network/Create new network/Transform/Line values/Set all values to 1
Network/Create new network/Transform/Add/Vertex labels/Default [yes]
Network/Create Partition/Degree/All
Partition/Binarize [1]             % Leaves
Partition/Copy to vector
Operations/Network+Vector/Network*Vector [1]    % r
Vector/Make Partition/by Intervals/Selected Thresholds [0.5]
Partition/Binarize [2]             % Roots
Partition/Copy to Vector
Operations/Network+Vector/Network*Vector [1]
select r as the second vector
Vectors/Add (First+Second)
Vector/Make Partition/by Intervals/Selected Thresholds [0.5]
Partition/Binarize Partition [1]   % Outsiders
Partition/Copy to Vector
Operations/Network+Vector/Network*Vector [1]
Vector/Make Partition/by Intervals/Selected Thresholds [0.5]
Partition/Binarize Partition [2]   % OutNeighbors
select Outsiders as the first and second partition
Partitions/Add (First+Second)
select  OutNeighbors as the second partition
Partitions/Add (First+Second)
Operations/Network+Partition/Extract [1-*]
Operations/Network+Partition/Transform/Remove lines/Inside clusters [1]
```

This sequence is saved as the macro `makePic`. It expects the US Airports network as the current network. To draw the network we select the last network and the last partition and run:

```
Draw/Network+First partition
Layout/Energy/Kamada-Kawai/Separate components
```

To improve the obtained picture some manual editing is needed.

# C Proofs

**Proposition 1**

(a) $0 \subseteq N[S] \subseteq V$ .

(b) $N[V] = V$.

(c) $W(S) = 1 \Leftrightarrow |N[S]| = n \Leftrightarrow N[S] = V \Leftrightarrow S$ is a dominanting set.

(d) $N(S_1) \subseteq N(S_2)$.

(e) trivial

**Proposition 2**

(a) in the paper.

(b) $N(V) \setminus V = \emptyset$.

(c) $S$ minimum dominating set $\Rightarrow |S| = \gamma \wedge N[S] = V$
$|N[S] \setminus S| = |V \setminus S| = |V| - |S| = n - \gamma \qquad \Rightarrow \qquad W^*(S) = 1$

(d) $S \subseteq N[S] \Rightarrow |N[S] \setminus S| = |N[S]| - |S|$
$W^*(S_1) - W^*(S_2) = \frac{1}{n-\gamma}(|N[S_1] - |S_1| - |N[S_2] + |S_2|) = \frac{1}{n-\gamma}(|S_2| - |S_1|) > 0.$

**Proposition 3** $k \leq \gamma$

(a) trivial.

(b) $W_{k_1}(S) = \frac{|N(S)\setminus S|}{n-k_1} < \frac{|N(S)\setminus S|}{n-k_2} = W_{k_2}(S)$ .

(c) from Prop 3.b and Prop 2.a: $W_k(S) < W_\gamma(S) = W^*(S) \leq 1$ .

(d) trivial.

(e) see Prop 2.d.