Modularity for community detection: history, perspectives and open issues

Vincenzo Nicosia [1]

[1]Dipartimento di Ingegneria Informatica e delle Telecomunicazioni
Università di Catania – Italy

March 12 2008 – Dipartimento di Matematica e Informatica – Catania (IT)

Outline

Community Definition

- Studies of complex networks revealed some interesting properties which can be exploited in communication and information networks:
  - Small–world effect
  - Self–organisation and self–adaption
  - Arising of community structure
- Small–world effect and Self–* properties have been exploited to optimise routing strategies and to increase network robustness
- On the other hand, Community structures have been deeply studied by physicians and mathematicians, but are quite unknown to researchers in the ITC field.

Community Definition

- Studies of complex networks revealed some interesting properties which can be exploited in communication and information networks:
  - Small–world effect
  - Self–organisation and self–adaption
  - Arising of community structure
- Small–world effect and Self–* properties have been exploited to optimise routing strategies and to increase network robustness
- On the other hand, Community structures have been deeply studied by physicians and mathematicians, but are quite unknown to researchers in the ITC field.

Community Definition

- Studies of complex networks revealed some interesting properties which can be exploited in communication and information networks:
  - Small–world effect
  - Self–organisation and self–adaption
  - Arising of community structure
- Small–world effect and Self–* properties have been exploited to optimise routing strategies and to increase network robustness
- On the other hand, Community structures have been deeply studied by physicians and mathematicians, but are quite unknown to researchers in the ITC field.

Community Definition

- Studies of complex networks revealed some interesting properties which can be exploited in communication and information networks:
  - Small–world effect
  - Self–organisation and self–adaption
  - Arising of community structure
- Small–world effect and Self–* properties have been exploited to optimise routing strategies and to increase network robustness
- On the other hand, Community structures have been deeply studied by physicians and mathematicians, but are quite unknown to researchers in the ITC field.

Communities in real world

- A community can be simply defined as a group of people which know each other and share interests and knowledge or collaborate to reach a given target
- Players of the same chess–club, classmates, members of a GNU/Linux User Group, students of engineering are examples of real–world communities....
- ....but also computers in the same LAN, people sharing rock music through eMule, programmers writing code for the same project are good examples of community structures
- Studying community structures in real–world networks can help to better understand and exploit community structures in information and communication networks
- Information about communities can be used to optimise message routing, to avoid waste of bandwidth and to speed–up collaboration and interaction in very large networks

Communities in real world

- A community can be simply defined as a group of people which know each other and share interests and knowledge or collaborate to reach a given target
- Players of the same chess–club, classmates, members of a GNU/Linux User Group, students of engineering are examples of real–world communities....
- ....but also computers in the same LAN, people sharing rock music through eMule, programmers writing code for the same project are good examples of community structures
- Studying community structures in real–world networks can help to better understand and exploit community structures in information and communication networks
- Information about communities can be used to optimise message routing, to avoid waste of bandwidth and to speed–up collaboration and interaction in very large networks

Communities in real world

- A community can be simply defined as a group of people which know each other and share interests and knowledge or collaborate to reach a given target
- Players of the same chess–club, classmates, members of a GNU/Linux User Group, students of engineering are examples of real–world communities....
- ....but also computers in the same LAN, people sharing rock music through eMule, programmers writing code for the same project are good examples of community structures
- Studying community structures in real–world networks can help to better understand and exploit community structures in information and communication networks
- Information about communities can be used to optimise message routing, to avoid waste of bandwidth and to speed–up collaboration and interaction in very large networks

Communities in real world

- A community can be simply defined as a group of people which know each other and share interests and knowledge or collaborate to reach a given target
- Players of the same chess–club, classmates, members of a GNU/Linux User Group, students of engineering are examples of real–world communities....
- ....but also computers in the same LAN, people sharing rock music through eMule, programmers writing code for the same project are good examples of community structures
- Studying community structures in real–world networks can help to better understand and exploit community structures in information and communication networks
- Information about communities can be used to optimise message routing, to avoid waste of bandwidth and to speed–up collaboration and interaction in very large networks

Communities in real world

- A community can be simply defined as a group of people which know each other and share interests and knowledge or collaborate to reach a given target
- Players of the same chess–club, classmates, members of a GNU/Linux User Group, students of engineering are examples of real–world communities....
- ....but also computers in the same LAN, people sharing rock music through eMule, programmers writing code for the same project are good examples of community structures
- Studying community structures in real–world networks can help to better understand and exploit community structures in information and communication networks
- Information about communities can be used to optimise message routing, to avoid waste of bandwidth and to speed–up collaboration and interaction in very large networks

Communities in real world

- A community can be simply defined as a group of people which know each other and share interests and knowledge or collaborate to reach a given target
- Players of the same chess–club, classmates, members of a GNU/Linux User Group, students of engineering are examples of real–world communities....
- ....but also computers in the same LAN, people sharing rock music through eMule, programmers writing code for the same project are good examples of community structures
- Studying community structures in real–world networks can help to better understand and exploit community structures in information and communication networks
- Information about communities can be used to optimise message routing, to avoid waste of bandwidth and to speed–up collaboration and interaction in very large networks

A community is....

- A sharp analytical definition of what a community is cannot be easily formulated. For this reason, many definitions of "community" have been proposed in the last few years
- One of the most widely used and accepted defines a community as "A group of nodes of a graph which are more strongly connected to each other than with other nodes in the same graph"
- Or as "A group of nodes of a graph which are more strongly connected to each other than expected in a corresponding random graph "
- Even if these definitions seem simple and clear, it is not easy to derive an algorithm to decide if a graph has communities and, in that case, to find which nodes belong to each community
- Note that the "community problem" has nothing in common with classical "clustering" problems. We're not looking for an optimal cut of a graph or for a min–flux decomposition....

A community is....

- A sharp analytical definition of what a community is cannot be easily formulated. For this reason, many definitions of "community" have been proposed in the last few years
- One of the most widely used and accepted defines a community as "A group of nodes of a graph which are more strongly connected to each other than with other nodes in the same graph"
- Or as "A group of nodes of a graph which are more strongly connected to each other than expected in a corresponding random graph "
- Even if these definitions seem simple and clear, it is not easy to derive an algorithm to decide if a graph has communities and, in that case, to find which nodes belong to each community
- Note that the "community problem" has nothing in common with classical "clustering" problems. We're not looking for an optimal cut of a graph or for a min–flux decomposition....

A community is....

- A sharp analytical definition of what a community is cannot be easily formulated. For this reason, many definitions of "community" have been proposed in the last few years
- One of the most widely used and accepted defines a community as "A group of nodes of a graph which are more strongly connected to each other than with other nodes in the same graph"
- Or as "A group of nodes of a graph which are more strongly connected to each other than expected in a corresponding random graph "
- Even if these definitions seem simple and clear, it is not easy to derive an algorithm to decide if a graph has communities and, in that case, to find which nodes belong to each community
- Note that the "community problem" has nothing in common with classical "clustering" problems. We're not looking for an optimal cut of a graph or for a min–flux decomposition....

A community is....

- A sharp analytical definition of what a community is cannot be easily formulated. For this reason, many definitions of "community" have been proposed in the last few years
- One of the most widely used and accepted defines a community as "A group of nodes of a graph which are more strongly connected to each other than with other nodes in the same graph"
- Or as "A group of nodes of a graph which are more strongly connected to each other than expected in a corresponding random graph "
- Even if these definitions seem simple and clear, it is not easy to derive an algorithm to decide if a graph has communities and, in that case, to find which nodes belong to each community
- Note that the "community problem" has nothing in common with classical "clustering" problems. We're not looking for an optimal cut of a graph or for a min–flux decomposition....

A community is....

- A sharp analytical definition of what a community is cannot be easily formulated. For this reason, many definitions of "community" have been proposed in the last few years
- One of the most widely used and accepted defines a community as "A group of nodes of a graph which are more strongly connected to each other than with other nodes in the same graph"
- Or as "A group of nodes of a graph which are more strongly connected to each other than expected in a corresponding random graph "
- Even if these definitions seem simple and clear, it is not easy to derive an algorithm to decide if a graph has communities and, in that case, to find which nodes belong to each community
- Note that the "community problem" has nothing in common with classical "clustering" problems. We're not looking for an optimal cut of a graph or for a min–flux decomposition....

A community is....

- A sharp analytical definition of what a community is cannot be easily formulated. For this reason, many definitions of "community" have been proposed in the last few years
- One of the most widely used and accepted defines a community as "A group of nodes of a graph which are more strongly connected to each other than with other nodes in the same graph"
- Or as "A group of nodes of a graph which are more strongly connected to each other than expected in a corresponding random graph "
- Even if these definitions seem simple and clear, it is not easy to derive an algorithm to decide if a graph has communities and, in that case, to find which nodes belong to each community
- Note that the "community problem" has nothing in common with classical "clustering" problems. We're not looking for an optimal cut of a graph or for a min–flux decomposition....

Finding communities: All of us agree.....

Finding communities: All of us agree.....

Finding communities: algorithms

Many different methods have been proposed for community identification. Danon et al. identify the following classes:

- Link removal methods (centrality–based)
- Agglomerative methods (e.g. Hierarchical clustering)
- Modularity optimization methods (a lot of them!)
- Spectral methods
- Other methods

We'll focus on modularity–based methods

Modularity and community structure

- A possible measure for evaluation of community decomposition is the so–called "Modularity"
- Given an undirected graph $G(E, V)$, where each node is assigned to one of C possible communities, the modularity of the decomposition is defined as (Newman):

$$Q = \frac{1}{2m} \sum_{i,j \in V} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

where:

- A_{ij} are the elements of the adjacency matrix of $G(E, V)$
- k_i is the out-degree of node i
- $m = |E|$
- $\delta(c_i, c_j)$ is equal to 1 if i and j belong to the same community, and is equal to 0 otherwise

Modularity and community structure

- A possible measure for evaluation of community decomposition is the so–called “Modularity”
- Given an undirected graph $G(E, V)$, where each node is assigned to one of C possible communities, the modularity of the decomposition is defined as (Newman):

$$Q = \frac{1}{2m} \sum_{i,j \in V} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

where:

  - A_{ij} are the elements of the adjacency matrix of $G(E, V)$
  - k_i is the out-degree of node i
  - $m = |E|$
  - $\delta(c_i, c_j)$ is equal to 1 if i and j belong to the same community, and is equal to 0 otherwise

Modularity and community structure

- A possible measure for evaluation of community decomposition is the so–called “Modularity”
- Given an undirected graph $G(E, V)$, where each node is assigned to one of C possible communities, the modularity of the decomposition is defined as (Newman):

$$Q = \frac{1}{2m} \sum_{i,j \in V} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)$$

where:

- A_{ij} are the elements of the adjacency matrix of $G(E, V)$
- k_i is the out-degree of node i
- $m = |E|$
- $\delta(c_i, c_j)$ is equal to 1 if i and j belong to the same community, and is equal to 0 otherwise

Modularity: what does it mean?

- The idea behind modularity definition is simple: a set of nodes form a community in the sense of modularity if the fraction of links inside the community is higher than expected in a network considered as “reference”(the so–called “null–model”)
- The term

$$\frac{1}{2m}\sum_{i,j\in V}[A_{ij}]\,\delta(c_i,c_j)$$

is the fraction of links connecting nodes which are in the same community, while

$$\frac{1}{2m}\sum_{i,j\in V}\left[\frac{k_i k_j}{2m}\right]\delta(c_i,c_j)$$

is the expected fraction of links connecting nodes in the same community in a random graph having the same degree distribution of the original graph

Modularity: what does it mean?

- The idea behind modularity definition is simple: a set of nodes form a community in the sense of modularity if the fraction of links inside the community is higher than expected in a network considered as “reference”(the so–called “null–model”)
- The term

$$\frac{1}{2m}\sum_{i,j\in V}[A_{ij}]\,\delta(c_i,c_j)$$

  is the fraction of links connecting nodes which are in the same community, while

$$\frac{1}{2m}\sum_{i,j\in V}\left[\frac{k_i k_j}{2m}\right]\delta(c_i,c_j)$$

  is the expected fraction of links connecting nodes in the same community in a random graph having the same degree distribution of the original graph

Notes about modularity

- Q is always lesser than 1, and equal to 0 only if all nodes are put in the same community
- High values of Q indicate a strong community structure
- Modularity is defined only for undirected graphs
- This definition allows a node to be put in just one community at a time, while nodes in real networks usually belong to many communities

Notes about modularity

- Q is always lesser than 1, and equal to 0 only if all nodes are put in the same community
- High values of Q indicate a strong community structure
- Modularity is defined only for undirected graphs
- This definition allows a node to be put in just one community at a time, while nodes in real networks usually belong to many communities

Notes about modularity

- Q is always lesser than 1, and equal to 0 only if all nodes are put in the same community
- High values of Q indicate a strong community structure
- Modularity is defined only for undirected graphs
- This definition allows a node to be put in just one community at a time, while nodes in real networks usually belong to many communities

Notes about modularity

- Q is always lesser than 1, and equal to 0 only if all nodes are put in the same community
- High values of Q indicate a strong community structure
- Modularity is defined only for undirected graphs
- This definition allows a node to be put in just one community at a time, while nodes in real networks usually belong to many communities

Notes about modularity

- Q is always lesser than 1, and equal to 0 only if all nodes are put in the same community
- High values of Q indicate a strong community structure
- Modularity is defined only for undirected graphs
- This definition allows a node to be put in just one community at a time, while nodes in real networks usually belong to many communities

Notes about modularity

- Q is always lesser than 1, and equal to 0 only if all nodes are put in the same community
- High values of Q indicate a strong community structure
- Modularity is defined only for undirected graphs
- This definition allows a node to be put in just one community at a time, while nodes in real networks usually belong to many communities

Extending modularity for directed graphs

- It is intuitively simple to extend the modularity definition for directed graphs (Arenas et al. 2007, Leicht – Newman 2007, Nicosia, Mangioni et al. 2008)
- It is necessary to consider a null–model where the probability of having a link between nodes i and j is the product of the probability that a link starts at i and of the probability that a link ends at j.
- So the extension of modularity for directed graphs can be written as:

$$Q_d = \frac{1}{m} \sum_{i,j \in V} \left[A_{ij} - \frac{k_i^{out} k_j^{in}}{m} \right] \delta(c_i, c_j)$$

- Note that:
  - Adjacency matrix is not symmetric for directed networks
  - The real fraction of links is referred to the real number of links in the graph (each link is counted just once!)

Extending modularity for directed graphs

- It is intuitively simple to extend the modularity definition for directed graphs (Arenas et al. 2007, Leicht – Newman 2007, Nicosia, Mangioni et al. 2008)
- It is necessary to consider a null–model where the probability of having a link between nodes i and j is the product of the probability that a link starts at i and of the probability that a link ends at j.
- So the extension of modularity for directed graphs can be written as:

$$Q_d = \frac{1}{m} \sum_{i,j \in V} \left[A_{ij} - \frac{k_i^{out} k_j^{in}}{m} \right] \delta(c_i, c_j)$$

- Note that:
  - Adjacency matrix is not symmetric for directed networks
  - The real fraction of links is referred to the real number of links in the graph (each link is counted just once!)

Extending modularity for directed graphs

- It is intuitively simple to extend the modularity definition for directed graphs (Arenas et al. 2007, Leicht – Newman 2007, Nicosia, Mangioni et al. 2008)
- It is necessary to consider a null–model where the probability of having a link between nodes i and j is the product of the probability that a link starts at i and of the probability that a link ends at j.
- So the extension of modularity for directed graphs can be written as:

$$Q_d = \frac{1}{m} \sum_{i,j \in V} \left[A_{ij} - \frac{k_i^{out} k_j^{in}}{m} \right] \delta(c_i, c_j)$$

- Note that:
  - Adjacency matrix is not symmetric for directed networks
  - The real fraction of links is referred to the real number of links in the graph (each link is counted just once!)

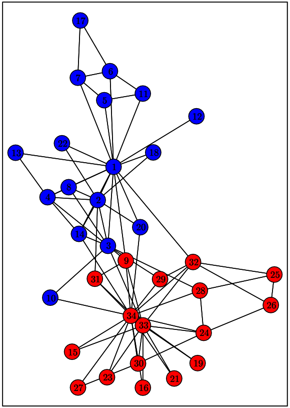
Extending modularity for directed graphs

- It is intuitively simple to extend the modularity definition for directed graphs (Arenas et al. 2007, Leicht – Newman 2007, Nicosia, Mangioni et al. 2008)
- It is necessary to consider a null–model where the probability of having a link between nodes i and j is the product of the probability that a link starts at i and of the probability that a link ends at j.
- So the extension of modularity for directed graphs can be written as:

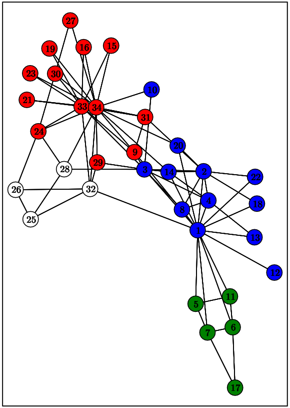
$$Q_d = \frac{1}{m} \sum_{i,j \in V} \left[A_{ij} - \frac{k_i^{out} k_j^{in}}{m} \right] \delta(c_i, c_j)$$

- Note that:
  - Adjacency matrix is not symmetric for directed networks
  - The real fraction of links is referred to the real number of links in the graph (each link is counted just once!)

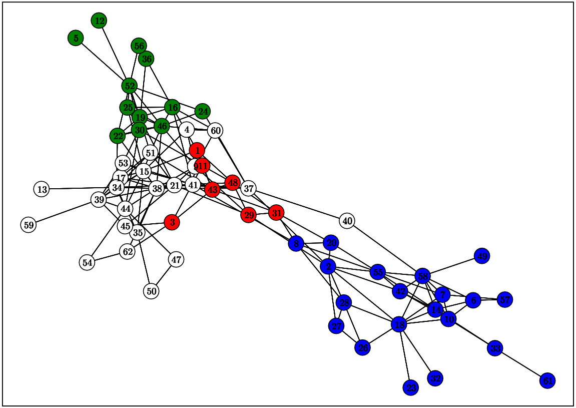
Modularity: The Zachary network (1)



Modularity: The Zachary network (2)



Modularity: The Dolphin network (1)



Drawbacks of modularity

Some drawbacks of modularity have been pointed out recently:

- It requires global knowledge of graph topology
- Many modularity optimization methods require a certain knowledge about the number of communities in the graph
- Modularity does not capture overlaps among communities in real networks (expecially in social and collaboration networks)
- Resolution limits (Fortunato, Barthelemy et al. 2006)

We recently proposed an extension to modularity which allows to take into account overlaps among communities in directed networks

Drawbacks of modularity

Some drawbacks of modularity have been pointed out recently:

- It requires global knowledge of graph topology
- Many modularity optimization methods require a certain knowledge about the number of communities in the graph
- Modularity does not capture overlaps among communities in real networks (expecially in social and collaboration networks)
- Resolution limits (Fortunato, Barthelemy et al. 2006)

We recently proposed an extension to modularity which allows to take into account overlaps among communities in directed networks

Drawbacks of modularity

Some drawbacks of modularity have been pointed out recently:

- It requires global knowledge of graph topology
- Many modularity optimization methods require a certain knowledge about the number of communities in the graph
- Modularity does not capture overlaps among communities in real networks (expecially in social and collaboration networks)
- Resolution limits (Fortunato, Barthelemy et al. 2006)

We recently proposed an extension to modularity which allows to take into account overlaps among communities in directed networks

Drawbacks of modularity

Some drawbacks of modularity have been pointed out recently:

- It requires global knowledge of graph topology
- Many modularity optimization methods require a certain knowledge about the number of communities in the graph
- Modularity does not capture overlaps among communities in real networks (expecially in social and collaboration networks)
- Resolution limits (Fortunato, Barthelemy et al. 2006)

We recently proposed an extension to modularity which allows to take into account overlaps among communities in directed networks

Drawbacks of modularity

Some drawbacks of modularity have been pointed out recently:

- It requires global knowledge of graph topology
- Many modularity optimization methods require a certain knowledge about the number of communities in the graph
- Modularity does not capture overlaps among communities in real networks (expecially in social and collaboration networks)
- Resolution limits (Fortunato, Barthelemy et al. 2006)

We recently proposed an extension to modularity which allows to take into account overlaps among communities in directed networks

Outline

Extension of modularity for overlapped communities

- Given a directed graph G(V,E) and a set of K overlapping communities of nodes of G, we can assing to each node i a vector of belonging coefficients $\alpha_{i,k}$
- $\alpha_{i,k}$ expresses how strongly node i belongs to community k, for each $k \in K$
- Without loss of generality, we can require that

$$0 \leq \alpha_{i,k} \leq 1 \forall i \in V, \quad \forall k \in K \tag{1}$$

and that

$$\sum_{k=1}^{|K|} \alpha_{i,k} = 1 \tag{2}$$

Extension of modularity for overlapped communities

- Given a directed graph G(V,E) and a set of K overlapping communities of nodes of G, we can assing to each node i a vector of belonging coefficients $\alpha_{i,k}$
- $\alpha_{i,k}$ expresses how strongly node i belongs to community k, for each $k \in K$
- Without loss of generality, we can require that

$$0 \leq \alpha_{i,k} \leq 1 \forall i \in V, \quad \forall k \in K \tag{1}$$

and that

$$\sum_{k=1}^{|K|} \alpha_{i,k} = 1 \tag{2}$$

Extension of modularity for overlapped communities

- Given a directed graph G(V,E) and a set of K overlapping communities of nodes of G, we can assing to each node i a vector of belonging coefficients $\alpha_{i,k}$
- $\alpha_{i,k}$ expresses how strongly node i belongs to community k, for each $k \in K$
- Without loss of generality, we can require that

$$0 \leq \alpha_{i,k} \leq 1 \forall i \in V, \quad \forall k \in K \tag{1}$$

and that

$$\sum_{k=1}^{|K|} \alpha_{i,k} = 1 \tag{2}$$

- We assign a belonging coefficient to community k to each edge l_{ij} connecting nodes i and j. We call it β_{lk}
- We imagine that β_{lk} is a certain function of the belonging coefficients of the nodes i and j that are connected by l_{ij}

$$\beta_{l,k} = \mathcal{F}(\alpha_{i,k}, \alpha_{j,k})$$

- We assign a belonging coefficient to community k to each edge l_{ij} connecting nodes i and j. We call it β_{lk}
- We imagine that β_{lk} is a certain function of the belonging coefficients of the nodes i and j that are connected by l_{ij}

$$\beta_{l,k} = \mathcal{F}(\alpha_{i,k}, \alpha_{j,k})$$

We can rewrite Newman's modularity distributing the factor $\delta(c_i, c_j)$ as follows:

$$Q_d = \frac{1}{m} \sum_{i,j \in V} \left[A_{ij} \delta(c_i, c_j) - \frac{k_i^{out} k_j^{in}}{m} \delta(c_i, c_j) \right] \tag{3}$$

In this case both the elements A_{ij} of the adjacency matrix and the probability of having a link between i and j in the null model are weighted by the belonging of i and j to the same community, i.e. $\delta(c_i, c_j)$

If we consider overlapped communities, the ontribute to the modularity of a community given by each edge is not sharp, and should be weighted by its belonging coefficient.

We can rewrite Newman's modularity distributing the factor $\delta(c_i, c_j)$ as follows:

$$Q_d = \frac{1}{m}\sum_{i,j\in V}\left[A_{ij}\delta(c_i,c_j) - \frac{k_i^{out}k_j^{in}}{m}\delta(c_i,c_j)\right] \quad (3)$$

In this case both the elements A_{ij} of the adjacency matrix and the probability of having a link between i and j in the null model are weighted by the belonging of i and j to the same community, i.e. $\delta(c_i, c_j)$

If we consider overlapped communities, the ontribute to the modularity of a community given by each edge is not sharp, and should be weighted by its belonging coefficient.

We can rewrite Newman's modularity distributing the factor $\delta(c_i, c_j)$ as follows:

$$Q_d = \frac{1}{m} \sum_{i,j \in V} \left[A_{ij} \delta(c_i, c_j) - \frac{k_i^{out} k_j^{in}}{m} \delta(c_i, c_j) \right] \quad (3)$$

In this case both the elements A_{ij} of the adjacency matrix and the probability of having a link between i and j in the null model are weighted by the belonging of i and j to the same community, i.e. $\delta(c_i, c_j)$

If we consider overlapped communities, the ontribute to the modularity of a community given by each edge is not sharp, and should be weighted by its belonging coefficient.

We can simply reformulate modularity, where $\delta(c_i, c_j)$ is substituted, respectively, by two different coefficients c_{ij} and d_{ij}, obtaining:

$$Q_{ov} = \frac{1}{m} \sum_{i,j \in V} \left[c_{ij} A_{ij} - d_{ij} \frac{k_i^{out} k_j^{in}}{m} \right] \quad (4)$$

It is also possible to put in evidence the contribution to modularity given by each community, so that we can rewrite the modularity as:

$$Q_{ov} = \frac{1}{m} \sum_{k \in K} \sum_{i,j \in V} \left[c_{ijk} A_{ij} - d_{ijk} \frac{k_i^{out} k_j^{in}}{m} \right] \quad (5)$$

c_{ijk} is the conribute to the modularity of community k due to $l_{i,j}$. As said above, it corresponds to the belonging coefficient of link $l_{i,j}$ to community k:

$$c_{ijk} = \beta_{l(i,j),k} = \beta_{l,k} = \mathcal{F}(\alpha_{i,k}, \alpha_{j,k}) \quad (6)$$

We can simply reformulate modularity, where $\delta(c_i, c_j)$ is substituted, respectively, by two different coefficients c_{ij} and d_{ij}, obtaining:

$$Q_{ov} = \frac{1}{m} \sum_{i,j \in V} \left[c_{ij} A_{ij} - d_{ij} \frac{k_i^{out} k_j^{in}}{m} \right] \quad (4)$$

It is also possible to put in evidence the contribution to modularity given by each community, so that we can rewrite the modularity as:

$$Q_{ov} = \frac{1}{m} \sum_{k \in K} \sum_{i,j \in V} \left[c_{ijk} A_{ij} - d_{ijk} \frac{k_i^{out} k_j^{in}}{m} \right] \quad (5)$$

c_{ijk} is the conribute to the modularity of community k due to $l_{i,j}$. As said above, it corresponds to the belonging coefficient of link $l_{i,j}$ to community k:

$$c_{ijk} = \beta_{l(i,j),k} = \beta_{l,k} = \mathcal{F}(\alpha_{i,k}, \alpha_{j,k}) \quad (6)$$

We can simply reformulate modularity, where $\delta(c_i, c_j)$ is substituted, respectively, by two different coefficients c_{ij} and d_{ij}, obtaining:

$$Q_{ov} = \frac{1}{m} \sum_{i,j \in V} \left[c_{ij} A_{ij} - d_{ij} \frac{k_i^{out} k_j^{in}}{m} \right] \tag{4}$$

It is also possible to put in evidence the contribution to modularity given by each community, so that we can rewrite the modularity as:

$$Q_{ov} = \frac{1}{m} \sum_{k \in K} \sum_{i,j \in V} \left[c_{ijk} A_{ij} - d_{ijk} \frac{k_i^{out} k_j^{in}}{m} \right] \tag{5}$$

c_{ijk} is the conribute to the modularity of community k due to $l_{i,j}$. As said above, it corresponds to the belonging coefficient of link $l_{i,j}$ to community k:

$$c_{ijk} = \beta_{l(i,j),k} = \beta_{l,k} = \mathcal{F}(\alpha_{i,k}, \alpha_{j,k}) \tag{6}$$

A neat definition of d_{ijk} is a bit more complicated, and requires a clear definition of the *null–model* to be used as reference.

given a graph $G(E,V)$ we choose as null–model a random graph corresponding to $G(E,V)$ where each node has out–degree and in–degree as in the original graph, and where the probability that a node i belongs to a given community k with a belonging factor $\alpha_{i,k}$ does not depend upon the probability that any other node j in the network does belong to the same community with $\alpha_{j,k}$.

This is equivalent to say that the expected belonging coefficient of any possible link $l(i,j)$ starting from a node into community k is simply the average of all possible belonging coefficients of l to k, so that:

$$\beta^{out}_{l(i,j),k} = \frac{\sum_{j\in V}\mathcal{F}(\alpha_{i,k},\alpha_{j,k})}{|V|} \tag{7}$$

and conversely:

$$\beta^{in}_{l(i,j),k} = \frac{\sum_{i\in V}\mathcal{F}(\alpha_{i,k},\alpha_{j,k})}{|V|} \tag{8}$$

A neat definition of d_{ijk} is a bit more complicated, and requires a clear definition of the *null–model* to be used as reference.
given a graph $G(E, V)$ we choose as null–model a random graph corresponding to $G(E, V)$ where each node has out–degree and in–degree as in the original graph, and where the probability that a node i belongs to a given community k with a belonging factor $\alpha_{i,k}$ does not depend upon the probability that any other node j in the network does belong to the same community with $\alpha_{j,k}$.
This is equivalent to say that the expected belonging coefficient of any possible link $l(i, j)$ starting from a node into community k is simply the average of all possible belonging coefficients of l to k, so that:

$$\beta^{out}_{l(i,j),k} = \frac{\sum_{j \in V} \mathcal{F}(\alpha_{i,k}, \alpha_{j,k})}{|V|} \tag{7}$$

and conversely:

$$\beta^{in}_{l(i,j),k} = \frac{\sum_{i \in V} \mathcal{F}(\alpha_{i,k}, \alpha_{j,k})}{|V|} \tag{8}$$

A neat definition of d_{ijk} is a bit more complicated, and requires a clear definition of the *null–model* to be used as reference.
given a graph $G(E, V)$ we choose as null–model a random graph corresponding to $G(E, V)$ where each node has out–degree and in–degree as in the original graph, and where the probability that a node i belongs to a given community k with a belonging factor $\alpha_{i,k}$ does not depend upon the probability that any other node j in the network does belong to the same community with $\alpha_{j,k}$.
This is equivalent to say that the expected belonging coefficient of any possible link $l(i, j)$ starting from a node into community k is simply the average of all possible belonging coefficients of l to k, so that:

$$\beta^{out}_{l(i,j),k} = \frac{\sum_{j \in V} \mathcal{F}(\alpha_{i,k}, \alpha_{j,k})}{|V|} \tag{7}$$

and conversely:

$$\beta^{in}_{l(i,j),k} = \frac{\sum_{i \in V} \mathcal{F}(\alpha_{i,k}, \alpha_{j,k})}{|V|} \tag{8}$$

As a consequence of reported considerations,

$$d_{ijk} = \beta^{out}_{l(i,j),k} \beta^{in}_{l(i,j),k}$$

and we can define modularity in the case of overlapped communities as:

$$Q_{ov} = \frac{1}{m} \sum_{k \in K} \sum_{i,j \in V} \left[\beta_{l(i,j),k} A_{ij} - \frac{\beta^{out}_{l(i,j),k} k^{out}_i \beta^{in}_{l(i,j),k} k^{in}_j}{m} \right] \tag{9}$$

We tested this definition with different $\mathcal{F}$s. Interesting results have been obtained using

$$\mathcal{F}(\alpha_{i,k}, \alpha_{j,k}) = \frac{1}{(1 + e^{-f(\alpha_{i,k})})(1 + e^{-f(\alpha_{j,k})})}$$

where f is a scaling function:

$$f(x) = 2kx - k \tag{10}$$

As a consequence of reported considerations,

$$d_{ijk} = \beta^{out}_{l(i,j),k}\beta^{in}_{l(i,j),k}$$

and we can define modularity in the case of overlapped communities as:

$$Q_{ov} = \frac{1}{m}\sum_{k\in K}\sum_{i,j\in V}\left[\beta_{l(i,j),k}A_{ij} - \frac{\beta^{out}_{l(i,j),k}k_i^{out}\beta^{in}_{l(i,j),k}k_j^{in}}{m}\right] \quad (9)$$

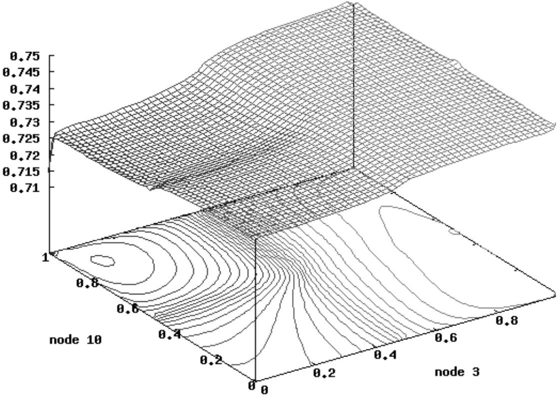
We tested this definition with different $\mathcal{F}$s. Interesting results have been obtained using

$$\mathcal{F}(\alpha_{i,k},\alpha_{j,k}) = \frac{1}{(1+e^{-f(\alpha_{i,k})})(1+e^{-f(\alpha_{j,k})})}$$

where f is a scaling function:

$$f(x) = 2kx - k \quad (10)$$

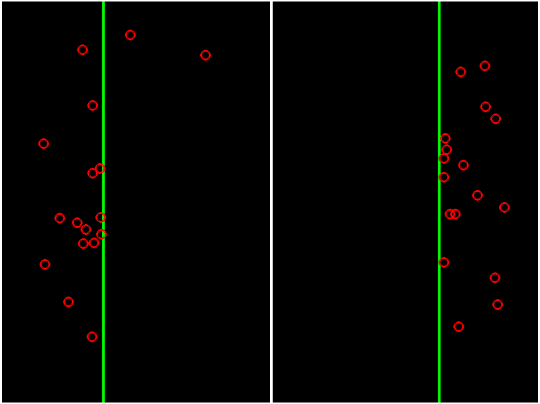
An example: the Zachary network



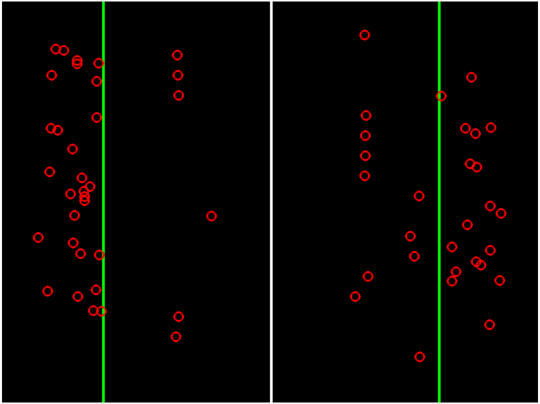
The highest value of modularity in the case of two communities is obtained when $\alpha_{3,1} = 0.81$ and $\alpha_{10,1} = 0.63$

Outline

Results: Zachary



Results: Engineering Students




"That's all Folks!"