

Multiple Optimal Solutions but Single Search: A Study of the Correlation Clustering Problem

Nejat Arinik

Rosa Figueiredo

Vincent Labatut

Laboratoire Informatique d'Avignon, University of Avignon, France.

Le Havre, Feb 21, 2019

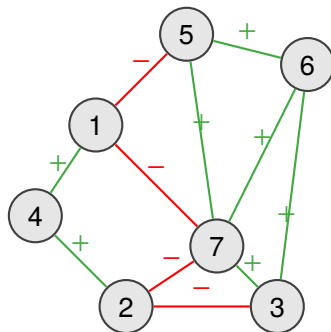
- 1 Context
- 2 Structural Balance and Signed Graph Partitioning
- 3 Methodology
- 4 Experiment
- 5 Conclusion

- **Motivation:** Multiple optimal solutions?
- **Objective:** Study of optimal solution space
- **Case Study:** Correlation Clustering (CC) problem
- **Input:** Signed graphs
- **Partitioning:** Structural Balance theory

[Arinik et al., 2019] Arinik, N. & Figueiredo, R. & Labatut, V. "*Multiple Partitioning of Multiplex Signed Networks: Application to European Parliament Votes*", Negative and Signed Tie Networks: Special Issue of Social Networks (*accepted*).

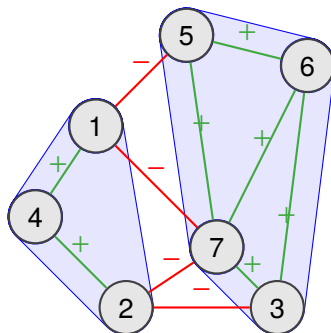
Structural Balance and Signed Graph Partitioning

- **Signed** graphs



Structural Balance and Signed Graph Partitioning

- **Signed** graphs
- **Structural Balance:** Partitioning into two [Heider, 1946] or more [Davis, 1964] mutually hostile subgroups each having internal solidarity → **Correlation Clustering (CC) problem**

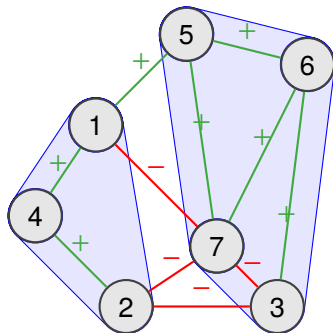


[Heider, 1946] Heider, F. "Attitudes and cognitive organization", Journal of Psychology, 21:107-112, 1946.

[Davis, 1964] Davis, J. "Clustering and structural balance in graphs", Human Relations, 20:181-187, 1967.

Structural Balance and Signed Graph Partitioning

- **Signed** graphs
- **Structural Balance:** Partitioning into two [Heider, 1946] or more [Davis, 1964] mutually hostile subgroups each having internal solidarity → **Correlation Clustering (CC) problem**
- Most real networks are not structurally balanced → need to measure graph **imbalance**



[Heider, 1946] Heider, F. "Attitudes and cognitive organization", Journal of Psychology, 21:107-112, 1946.

[Davis, 1964] Davis, J. "Clustering and structural balance in graphs", Human Relations, 20:181-187, 1967.

- Imbalance of a partition

$P = \{S_1, S_2, \dots, S_l\}$ of V

$$I(P) = \sum_{1 \leq i \leq l} \Omega^-(S_i, S_i) + \sum_{1 \leq i < j \leq l} \Omega^+(S_i, S_j).$$

where

$$\Omega^+(S_i, S_j) = \sum_{e \in E^+ \cap E[S_i:S_j]} w_e$$

$$\text{and } \Omega^-(S_i, S_j) = \sum_{e \in E^- \cap E[S_i:S_j]} w_e$$

Measuring imbalance - CC problem

- Imbalance of a partition

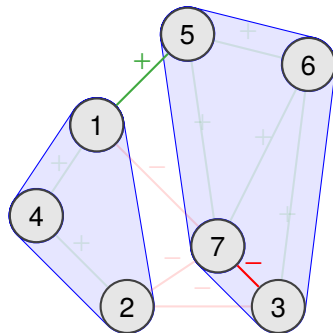
$P = \{S_1, S_2, \dots, S_l\}$ of V

$$I(P) = \sum_{1 \leq i \leq l} \Omega^-(S_i, S_i) + \sum_{1 \leq i < j \leq l} \Omega^+(S_i, S_j).$$

where

$$\Omega^+(S_i, S_j) = \sum_{e \in E^+ \cap E[S_i : S_j]} w_e$$

$$\text{and } \Omega^-(S_i, S_j) = \sum_{e \in E^- \cap E[S_i : S_j]} w_e$$



$$I(P) = 0 + 1 + 1 = 2$$

Measuring imbalance - CC problem

Definition

Consider a signed graph $G = (V, E, s)$ with a nonnegative weight for each $e \in E$. The **Correlation Clustering (CC) problem** is the problem of finding a partition P of V such that the imbalance $I(P)$ is minimized.

- Imbalance of a partition

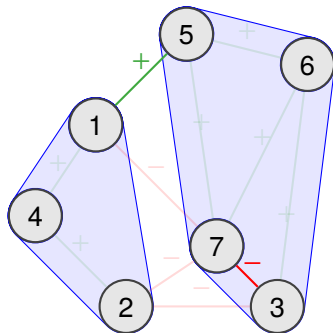
$$P = \{S_1, S_2, \dots, S_l\} \text{ of } V$$

$$I(P) = \sum_{1 \leq i \leq l} \Omega^-(S_i, S_i) + \sum_{1 \leq i < j \leq l} \Omega^+(S_i, S_j).$$

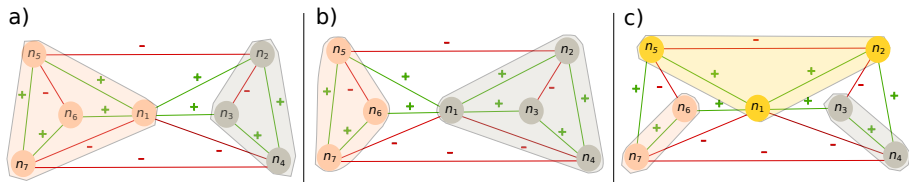
where

$$\Omega^+(S_i, S_j) = \sum_{e \in E^+ \cap E[S_i: S_j]} w_e$$

$$\text{and } \Omega^-(S_i, S_j) = \sum_{e \in E^- \cap E[S_i: S_j]} w_e$$

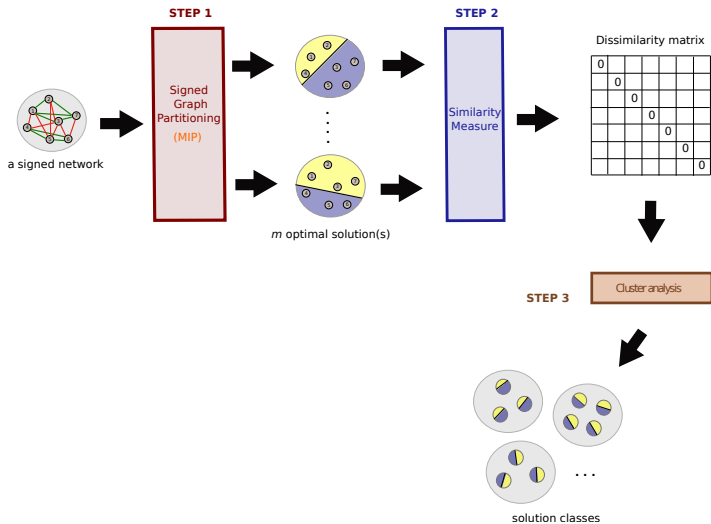


Example



$I(P) = 4$
22 optimal solutions

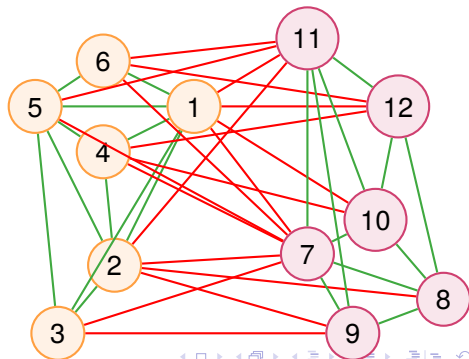
Methodology



Experiment

- Random (complete and unweighted) graph generation process
 - First: generate perfectly balanced random signed graphs \rightarrow graph size (n) and an initial number of cluster (k_{init})

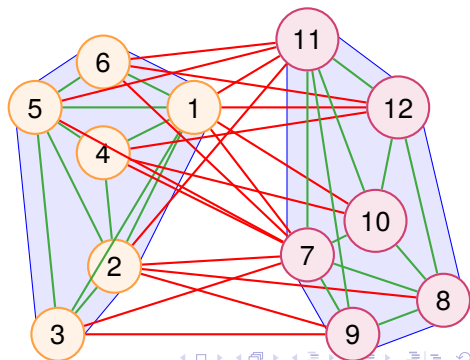
- $n = 12, k_{init} = 2$



Experiment

- Random (complete and unweighted) graph generation process
 - First: generate perfectly balanced random signed graphs \rightarrow graph size (n) and an initial number of cluster (k_{init})

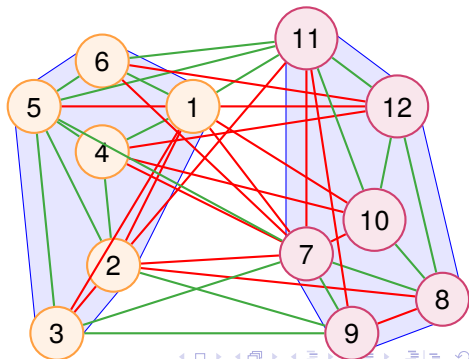
- $n = 12, k_{init} = 2$



Experiment

- Random (complete and unweighted) graph generation process
 - First: generate perfectly balanced random signed graphs \rightarrow graph size (n) and an initial number of cluster (k_{init})
 - Then: make graphs imbalanced \rightarrow proportion of misplaced links (q_{mispl})

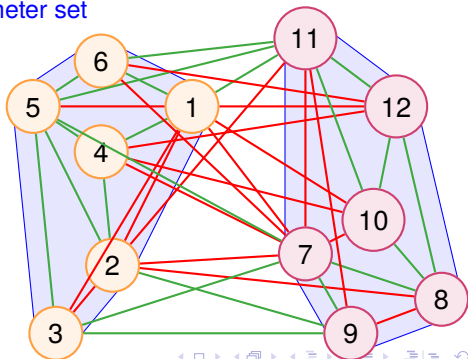
- $n = 12, k_{init} = 2$
- $q_{mispl} = 0.20$



Experiment

- Random (complete and unweighted) graph generation process
 - First: generate perfectly balanced random signed graphs \rightarrow graph size (n) and an initial number of cluster (k_{init})
 - Then: make graphs imbalanced \rightarrow proportion of misplaced links (q_{mispl})
- Input graphs
 - $k_{init} = 2, n \in \{16, 20, 24, 28, 32, 36\}, q_{mispl} \in \{0.05, 0.10, 0.15, \dots, 0.70\}$
 - $k_{init} = 4, n \in \{16, 20, 24, 28, 32, 36\}, q_{mispl} \in \{0.05, 0.10, 0.15, \dots, 0.40\}$
 - 10 instances for each parameter set

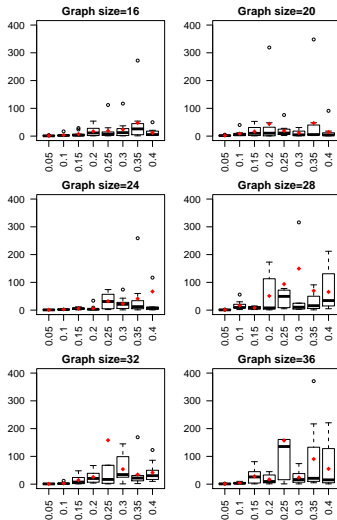
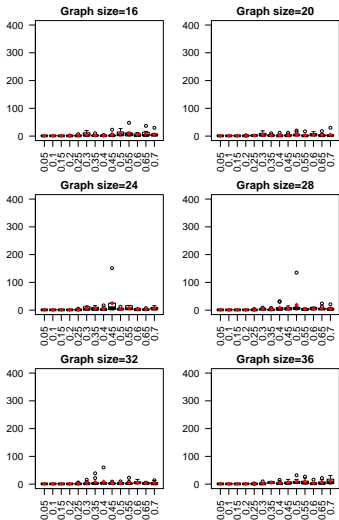
- $n = 12, k_{init} = 2$
- $q_{mispl} = 0.20$



Preliminary results: Number of optimal solutions

$k_{init} = 2$

$k_{init} = 4$

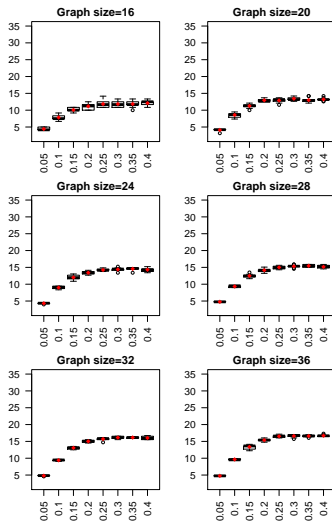
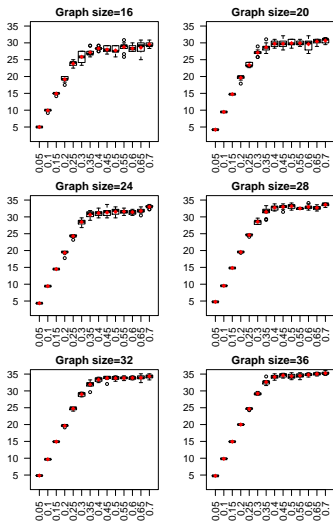


x-axis: q_{mispl} (Proportion of misplaced links based on k_{init})

Preliminary results: Graph Imbalance (%)

$k_{init} = 2$

$k_{init} = 4$

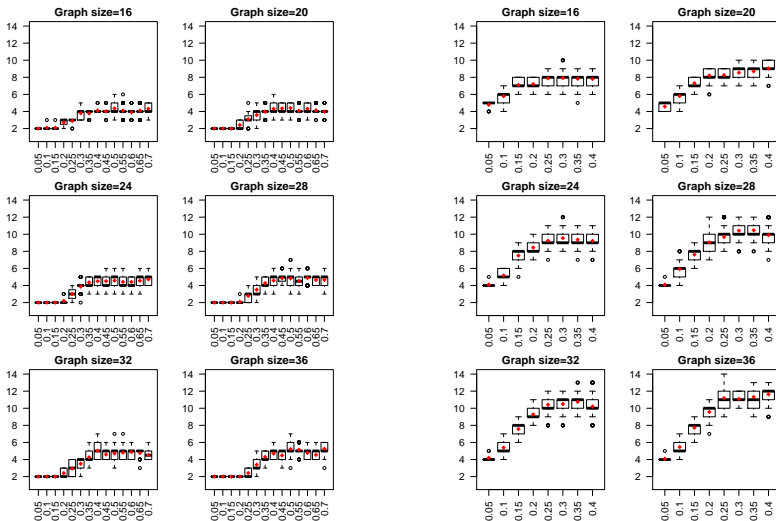


x-axis: q_{mispl} (Proportion of misplaced links based on k_{init})

Prel. results: Number of detected clusters (k) in solutions

$k_{init} = 2$

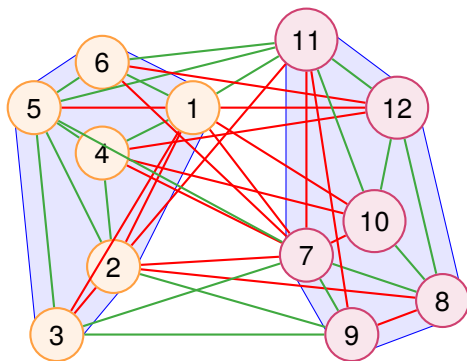
$k_{init} = 4$



x-axis: q_{mispl} (Proportion of misplaced links based on k_{init})

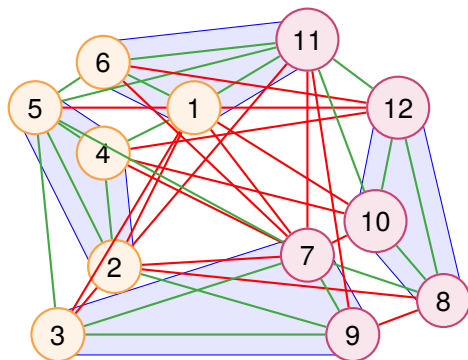
Number of detected clusters (k) in solutions

- $n = 12, k_{init} = 2$
- $q_{mispl} = 0.20$



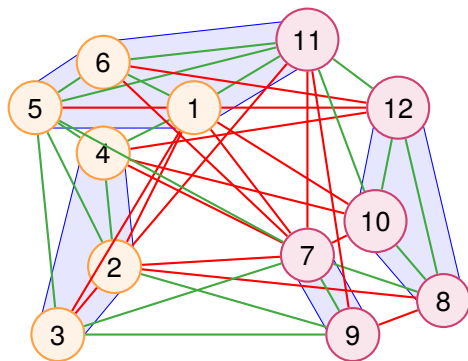
Number of detected clusters (k) in solutions

- $n = 12, k_{init} = 2$
- $q_{mispl} = 0.20$
- $k = 4$



Number of detected clusters (k) in solutions

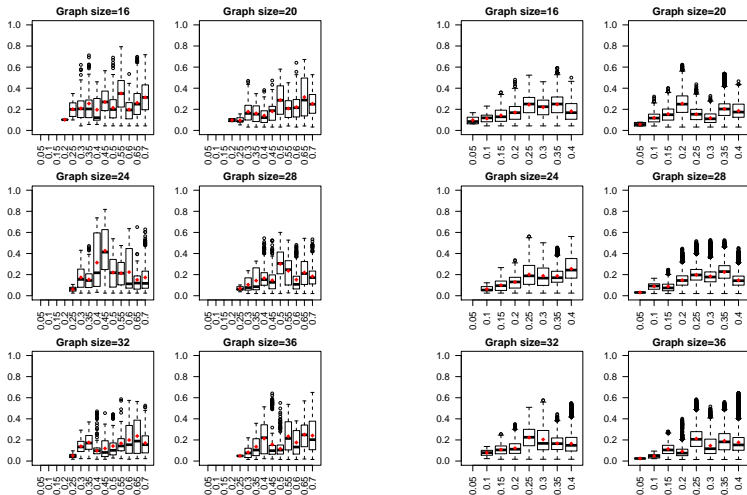
- $n = 12$, $k_{init} = 2$
- $q_{mispl} = 0.20$
- $k = 4$



Preliminary results: Dissimilarity of optimal solutions

$k_{init} = 2$

$k_{init} = 4$



x-axis: q_{mispl} (Proportion of misplaced links based on k_{init})

Detecting solution classes

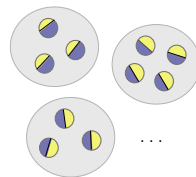
Dissimilarity matrix

0						
	0					
		0				
			0			
				0		
					0	
						0



Cluster analysis

STEP 3

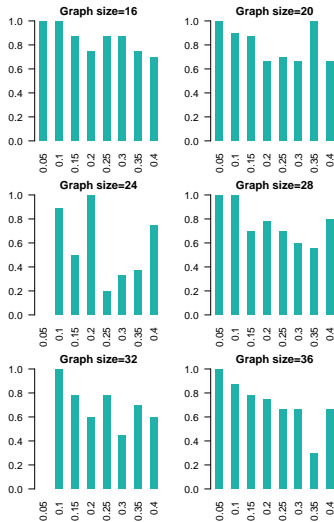
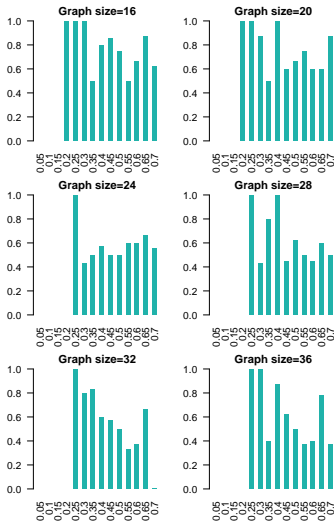


solution classes

Preliminary results: Proportion of single class case

$k_{init} = 2$

$k_{init} = 4$



x-axis: q_{mispl} (Proportion of misplaced links based on k_{init})

Conclusion & Further research

- Complete and unweighted signed networks
 - Many optimal solutions: %49 ($k_{init} = 2$) and %89 ($k_{init} = 4$) of the instances
 - Can be very different: %35 ($k_{init} = 2$) and %25 ($k_{init} = 4$) of the cases have different solution classes
-
- Study of operators
 - Analysis of heuristic methods

Thank you for your attention!

Contact Information:

Nejat ARINIK

nejat.arinik@univ-avignon.fr

Correlation-Clustering Problem - ILP formulation

$$x_{ij} = \begin{cases} 0 & \text{if vertex } i \text{ and } j \text{ are in a common set,} \\ 1 & \text{otherwise.} \end{cases}$$

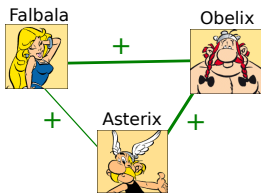
$$\text{minimize } \sum_{(i,j) \in E^-} w_{ij}(1 - x_{ij}) + \sum_{(i,j) \in E^+} w_{ij}x_{ij}$$

$$\text{subject to } x_{ip} + x_{pj} \geq x_{ij}, \quad \forall i, p, j \in V, \quad (1)$$

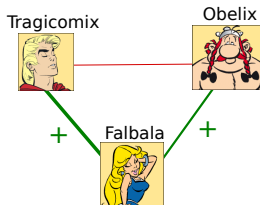
$$x_{ij} = x_{ji}, \quad \forall i, j \in V, \quad (2)$$

$$x_{ij} \in \{0, 1\}, \quad \forall i, j \in V. \quad (3)$$

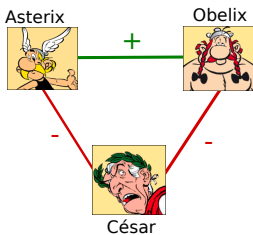
Structural Balance



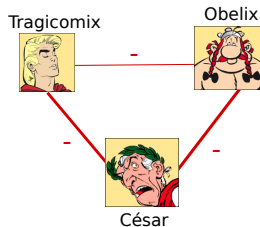
(a) Balanced



(b) Not balanced

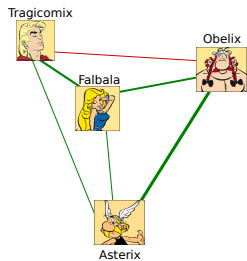


(c) Balanced

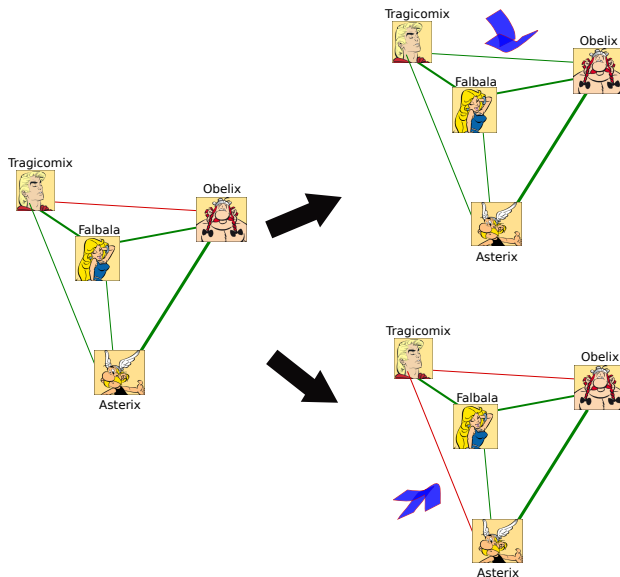


(d) Not balanced

Structural Balance

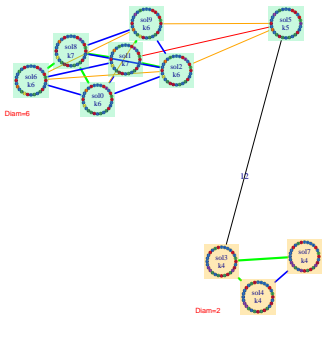


Structural Balance

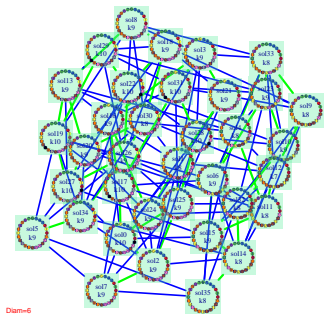


Preliminary results: Transition graph

$n = 28$, $k_{init} = 2$, $q_{mispl} = 0.50$, network=5

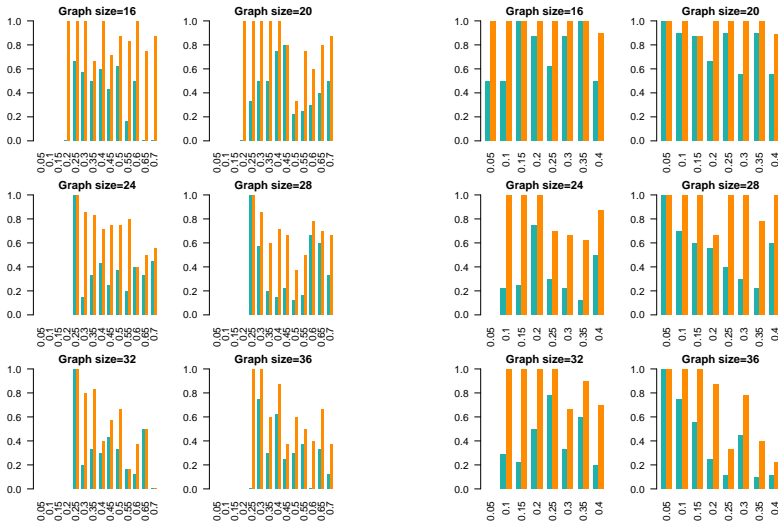


$n = 24$, $k_{init} = 4$, $q_{mispl} = 0.25$, network=7



Number of Connected Components in Transition Graphs

— Proportion of 1 component with 1-Edit — Proportion of 1 component with 1-Edit, 2-Edit, 3-Edit and 4-Edit



x-axis: q_{misl} (Proportion of misplaced links)