

# Network Analysis

*Group 7 Members:*

*Chen Gao*

*Qiang Wang*

*Toshan Dubhashi*

*Yidan Wang*

# Introduction & Data Preparation

- Basic Information :
  - Datasets Used - InfoSeek.xlsx, Collab.xlsx, Attributes.xlsx
  - Libraries Used - readxl, dplyr, igraph
  - Tools Used - R Studio, Gephi
- Data Preparation :
  - Excluded the ID Column from both InfoSeek and Collab dataset
  - Dichotomized the data - Keeping only links with strength value greater than 3
  - Binary Matrix formation - 1 signifies edge & 0 signifies absence of edge
  - Setting all the diagonal matrix elements to 0

# Analysis Focus

- Is there any relationship between the satisfaction level and the corresponding node attribute ?
- Who are the key players in the corporate based on their info seeking and collaborating behaviors ?
- What key factors affect the InfoSeek and Collab relationships among them and if one relationship can predict the other ?
- What kind of employees are more likely to collaborate and share work with other staff members ?

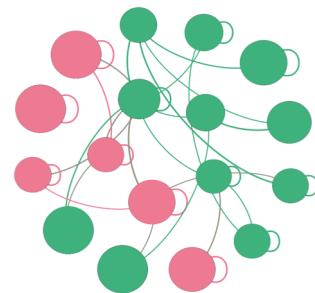
# Overall Satisfaction

Node Size : Satisfaction  
Node Color : Attribute  
Range: Top 15.6% or Bottom 15%

## Gender

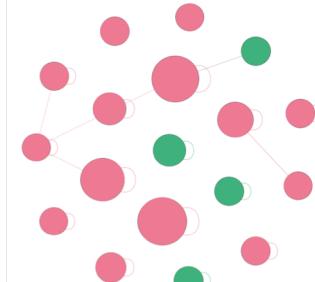
1: 55.05%

2: 44.95%



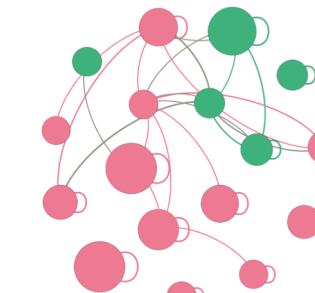
35%

64%



75%

25%



75%

25%

## Department:

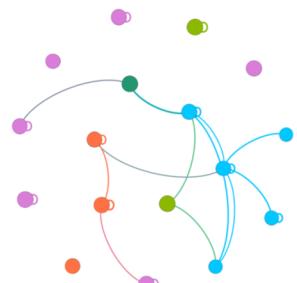
2 : 30.28%

4 : 25.69%

3 : 21.1%

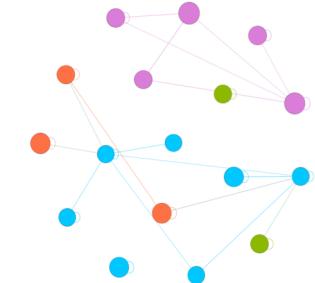
5 : 20.18%

1 : 2.75%



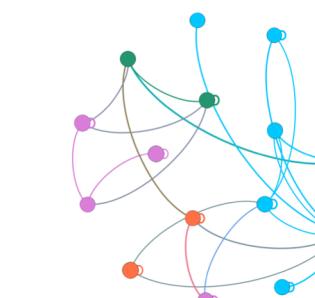
29%

11%



41%

12%



47%

0%

Individual

Group

Organization

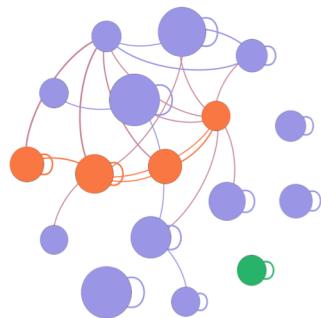
# Overall Satisfaction

## Tenure

1: 51.38%

2: 27.52%

3: 21.1%



70%

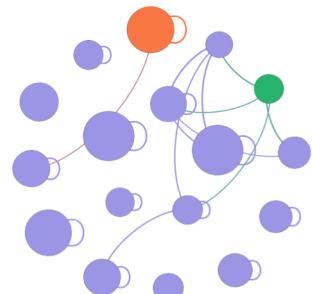
5.8%

## Management

2: 69.72%

1: 22.94%

3: 7.34%

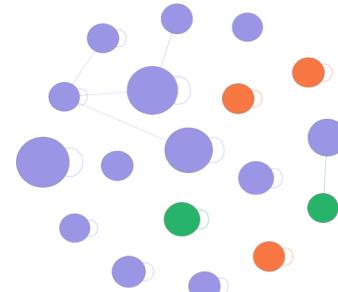


88.4%

5.8%

5.8%

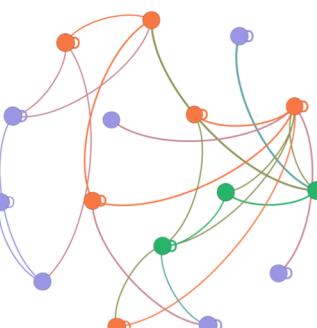
## Group



73%

2.16%

3.11%



41%

35%

17%

## Organization

Node Size : Satisfaction

Node Color : Attribute

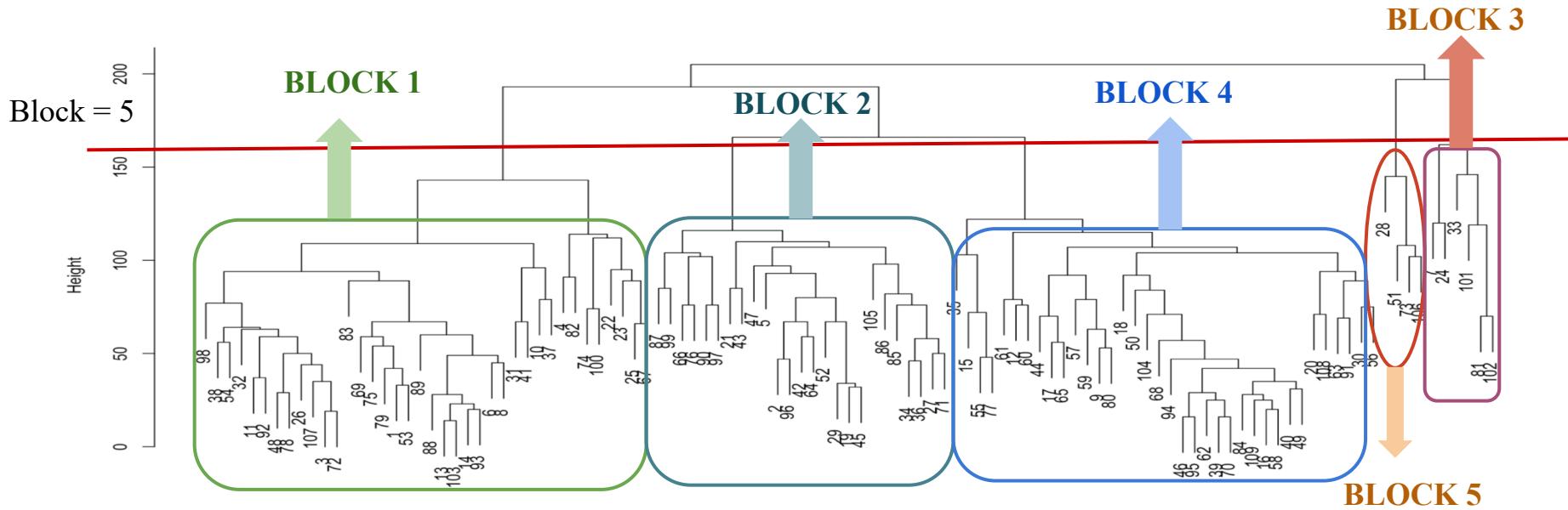
Range: Top 15.6% or Bottom 15%

# Role & Descriptive Analysis

- Block Model
- Indegree & Outdegree
- Degree Centrality, Closeness Centrality  
Betweenness Centrality, Coreness
- Gephi & Attributes

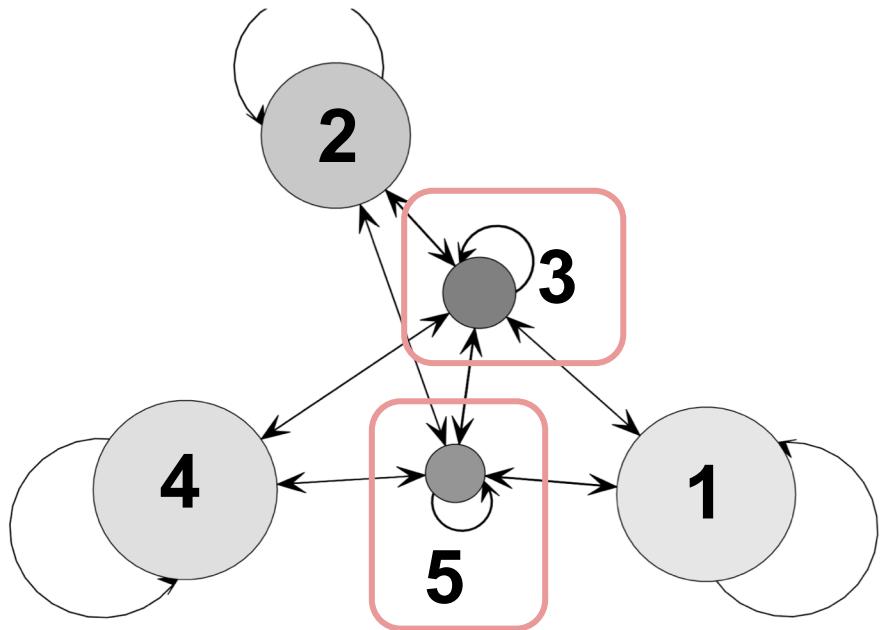


# Cluster Dendrogram

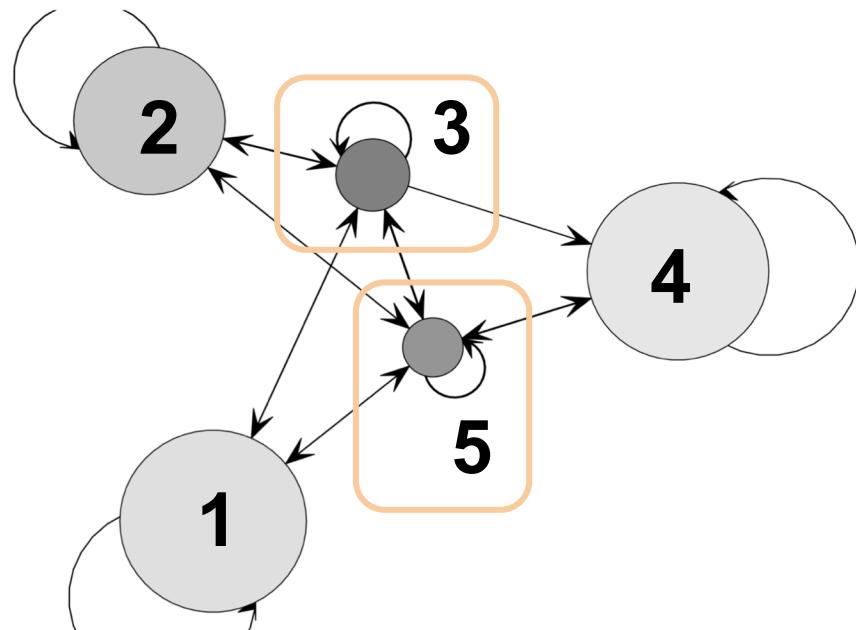


- BLOCK 5: D2R200, D5R136, D2R92, D2R58
- BLOCK 3: D4R189, D4R149, D1R187, D1R66, D4R53, D3R14

# Block Model



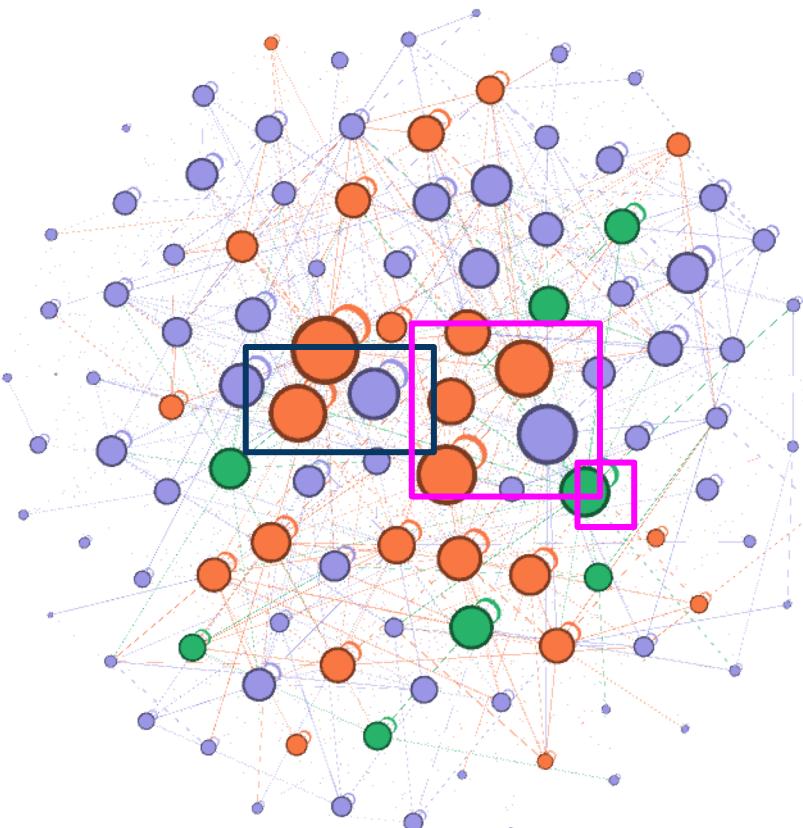
Info Seeking Behavior



Collaborating Behavior

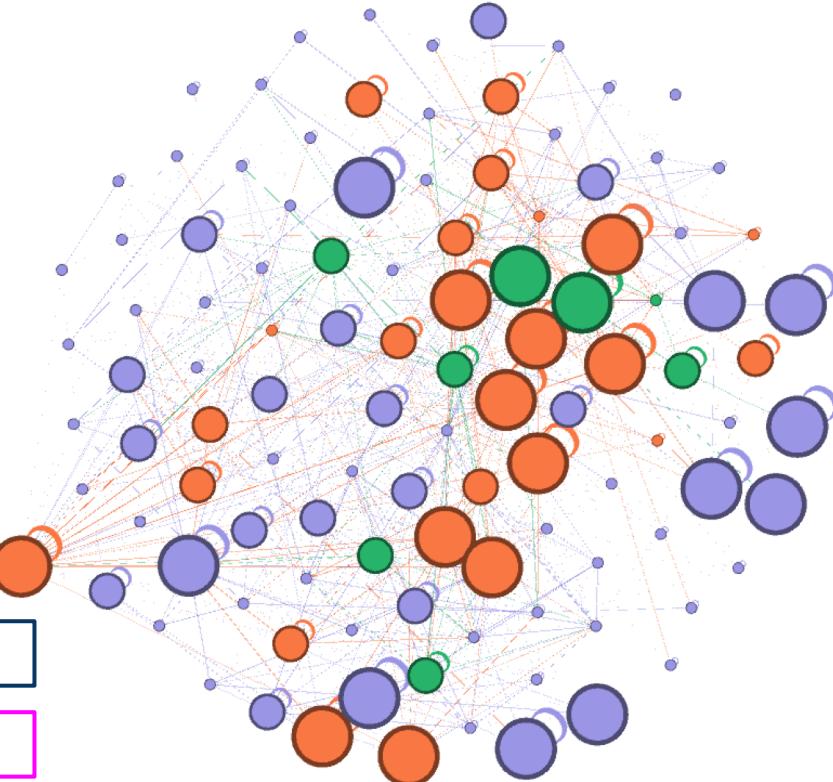
# BLOCK 5 & BLOCK 3

Node Size : Degree/ Tenure  
Node Color : Management Level



Info Seeking Behavior

BLOCK5  
BLOCK3



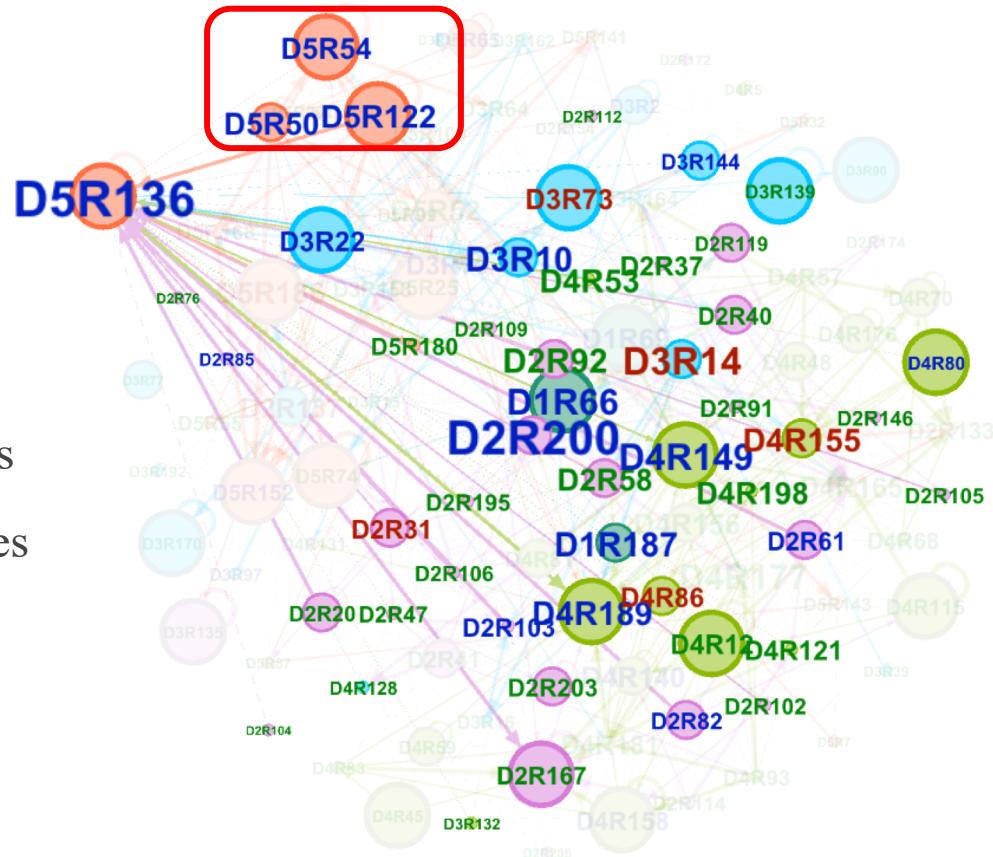
Collaborating Behavior

## BLOCK 5 & BLOCK 3: Ranks in Centrality

# Info Seeking Collaborating

# D5R136

- Long tenure
- High management level
- Connection with other departments
- High indegree and outdegree values
- Dominant position in D5



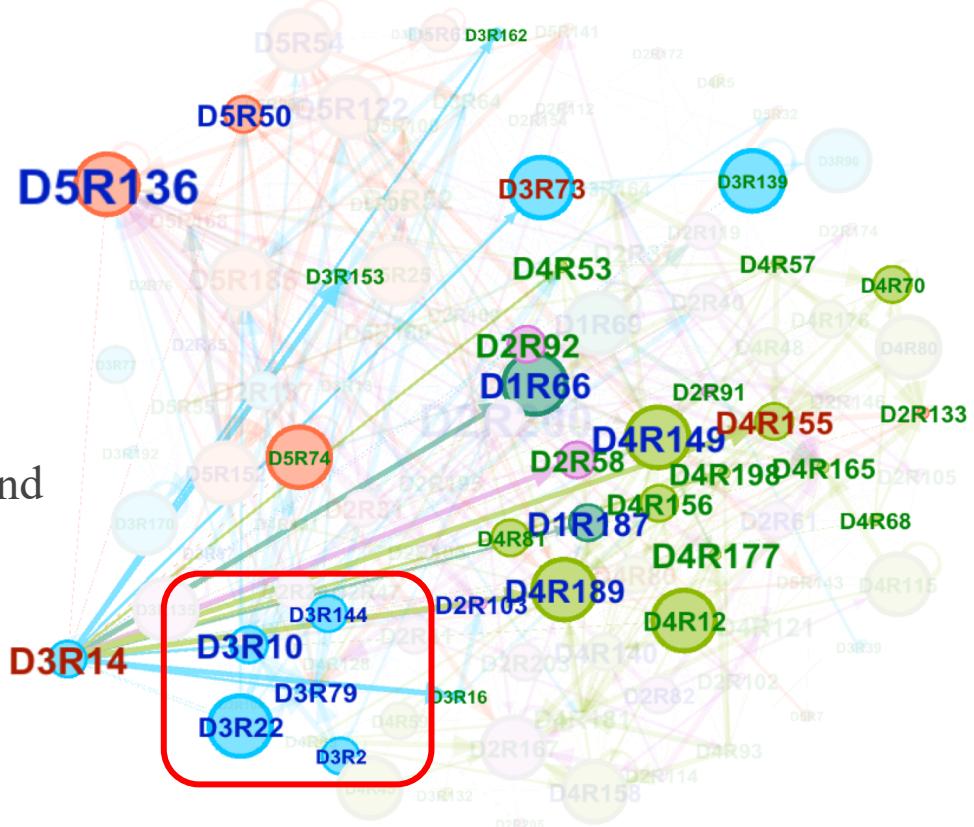
Node Size: Tenure  
Label Size: Degree

Node Color: Department  
Label Color: Management Level

Collaborating Behavior

# D3R14

- Short tenure
- Connection with other departments
- No dominant managers in D3
- Serve as the intermediary among D3 and other departments



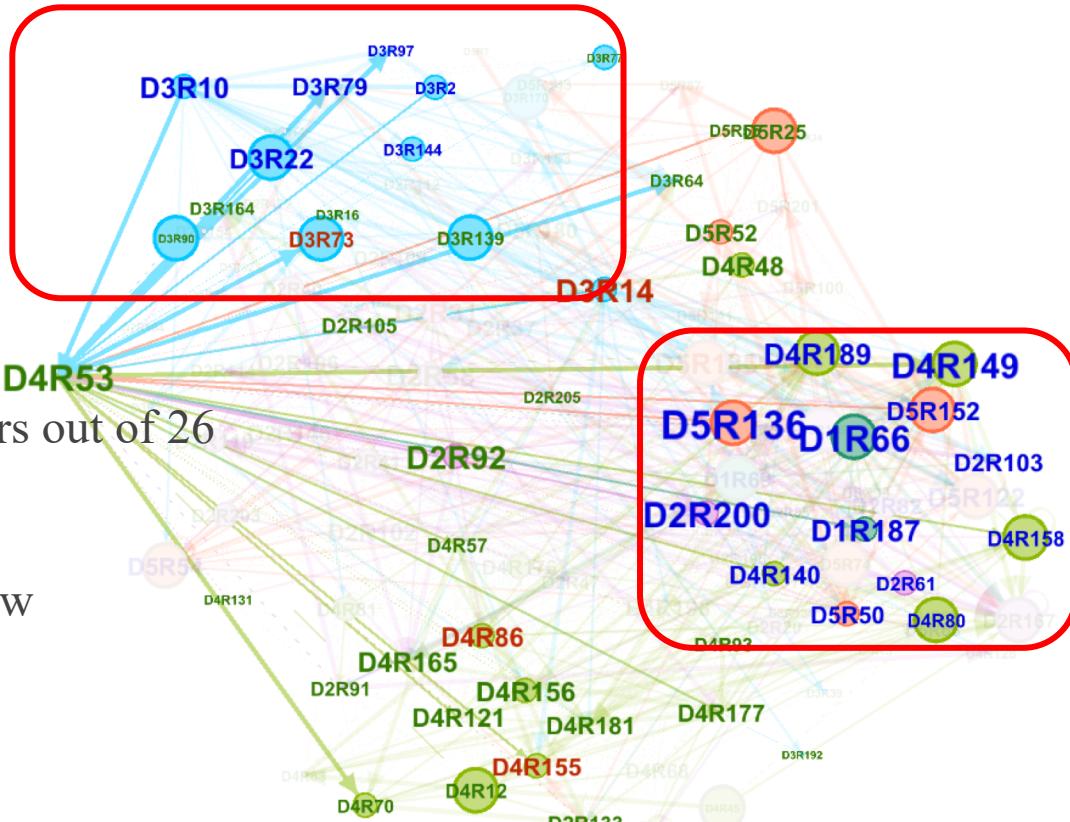
Node Size: Tenure  
Label Size: Degree

Node Color: Department  
Label Color: Management Level

Info Sabotaging Behavior

D4R53

- Shortest tenure
  - Lowest Management level 2
  - Highly connected with 20 managers out of 26
  - Significant relationships with D3,  
whose overall degree values are low



Node Size: Tenure  
Label Size: Degree

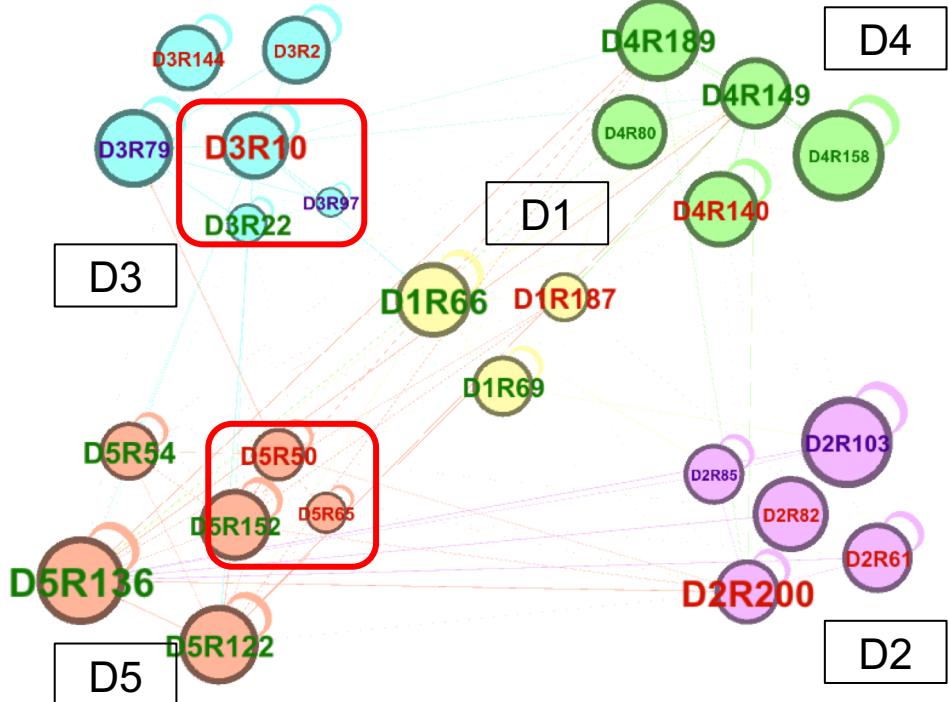
Node Color: Department  
Label Color: Management Level

# Info Seeking Behavior

# Satisfaction Level among Managers

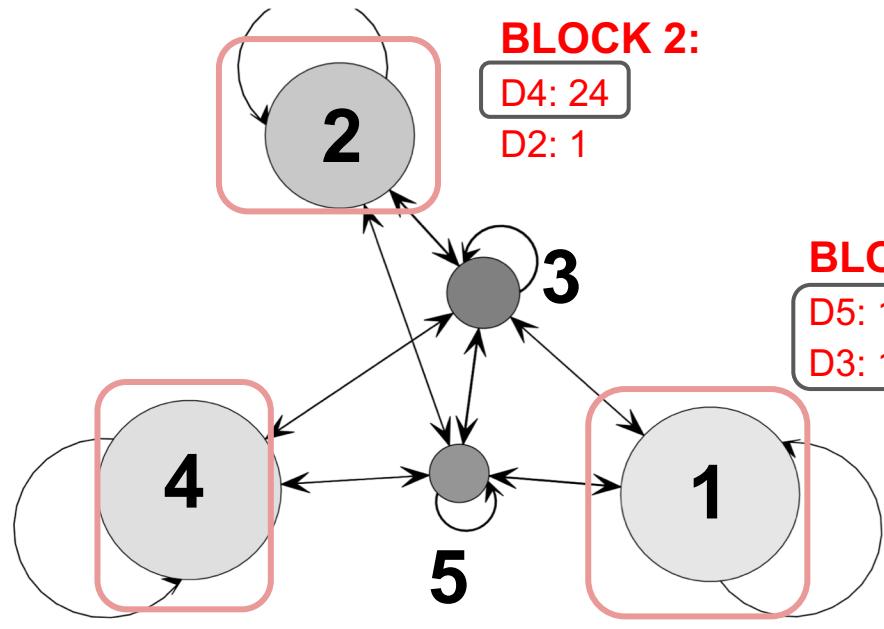
- Nodes with a smaller degree are of lower satisfaction
- A high proportion of managers in D3 and D5 are unsatisfied with group and organization productivity
- High satisfaction of individual productivity

Node Size: Satisfaction  
Label Size: Degree  
Node Color: Department  
Label Color: Gender



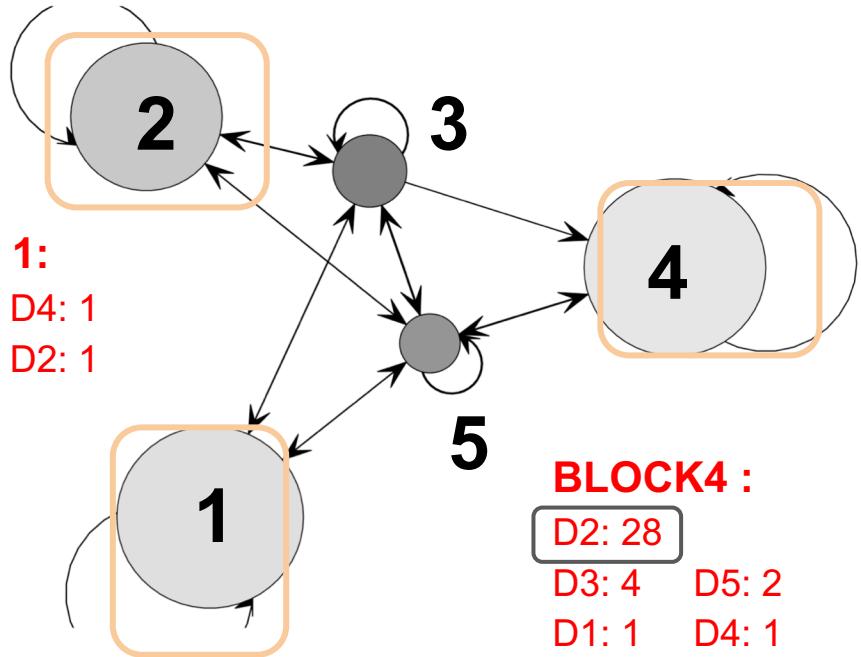
Satisfaction of Individual Productivity

# Block Model



**BLOCK 1:**

- D5: 18
- D3: 18



Info Seeking Behavior

Collaborating Behavior

# Statistical Analysis

- QAP
- Netlogit function performs a logistic regression of the network variable in y on the network variables in set x.
- ERGM
- ERGM can be used to understand whether a given observed network can be derived from characteristics of the network members

# QAP

Network Logit Model

Coefficients:

	Estimate	Exp(b)	Pr(<=b)	Pr(>=b)	Pr(>= b )
(intercept)	-2.477634	0.08394161	0	1	0
x1	5.003363	148.91304348	1	0	0

Odds Ratio  
Exp(b): 148.91

Goodness of Fit Statistics:

Null deviance: 16319.46 on 11772 degrees of freedom

Residual deviance: 6376.939 on 11770 degrees of freedom

Chi-Squared test of fit improvement:

9942.519 on 2 degrees of freedom, p-value 0

AIC: 6380.939 BIC: 6395.686

Pseudo-R^2 Measures:

$(D_n - D_r) / (D_n - D_r + df_n)$ : 0.4578742

$(D_n - D_r) / D_n$ : 0.6092432

Contingency Table (predicted (rows) x actual (cols)):

Actual			
		0	1
Predicted	0	8768	736
	1	168	2100

Confusion Matrix

Total Fraction Correct: 0.9232076

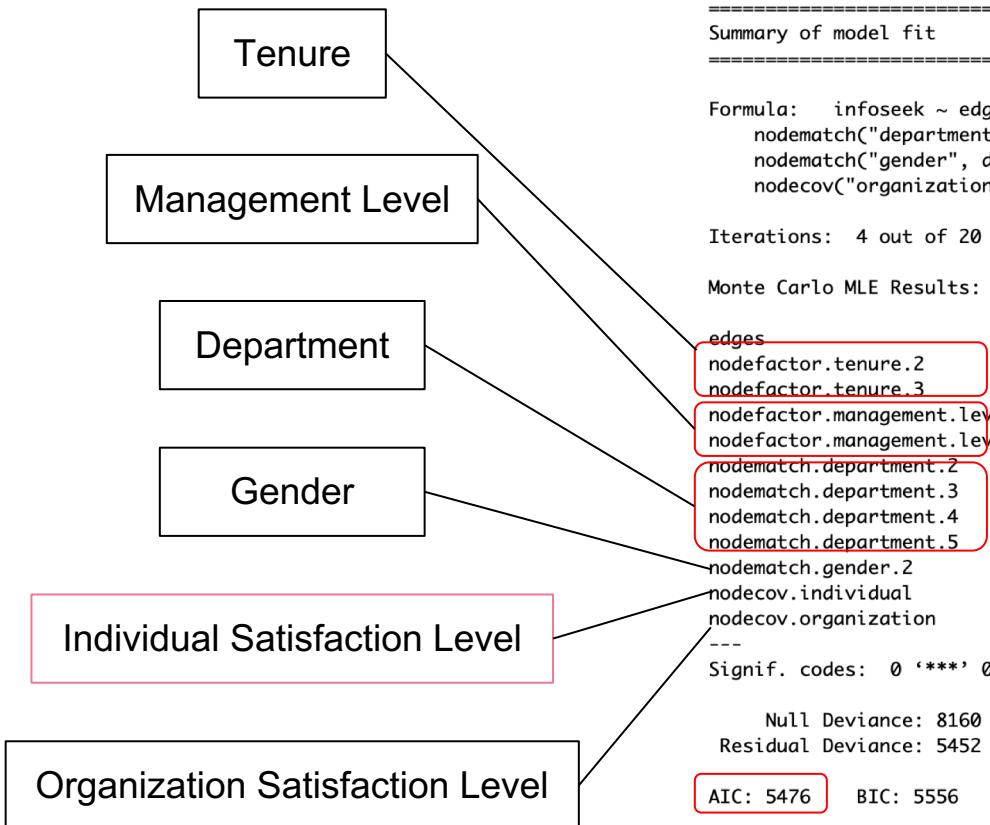
Fraction Predicted 1s Correct: 0.9259259

Fraction Predicted 0s Correct: 0.9225589

False Negative Rate: 0.2595205

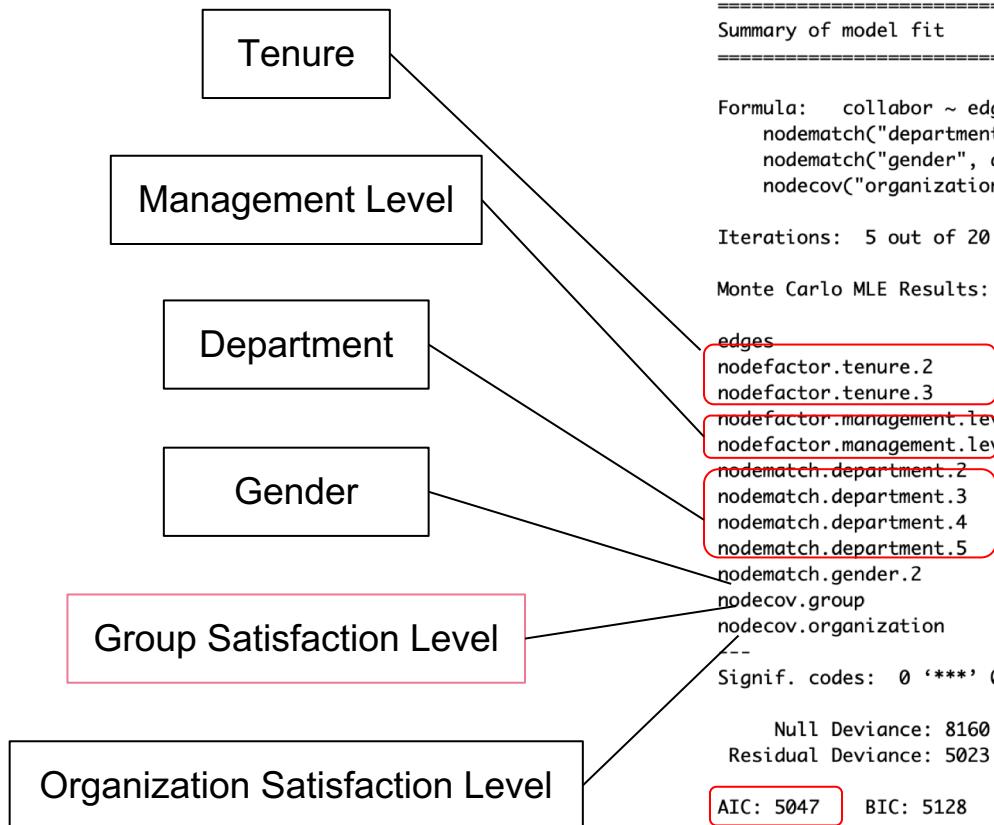
False Positive Rate: 0.01880036

# ERGM: Key Factors of Info Seeking



Null model AIC: 6502

# ERGM: Key Factors of Collaboration



Null model AIC: 5771

# SUMMARY

- QAP
- ERGM
  - Organization Satisfaction Level has a negative effect on info-seeking and collaboration relationship
  - Individual Satisfaction Level has a positive effect on info-seeking relationship
  - Group Satisfaction Level has a positive effect on collaboration relationship

