



# Network-based Visual Analysis of Tabular Data

Zhicheng Liu, Shamkant Navathe, John Stasko



# Tabular Data



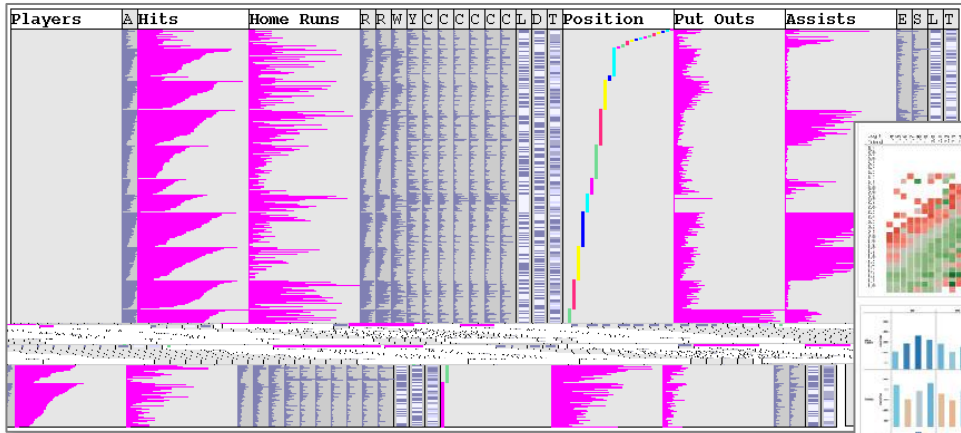
# Tabular Data

- Rows and columns
- Rows are **data cases**; columns are **attributes/dimensions**
- Attribute types
  - Quantitative (numbers)
  - Ordinal (e.g. small, medium, large)
  - Nominal (names, categories)

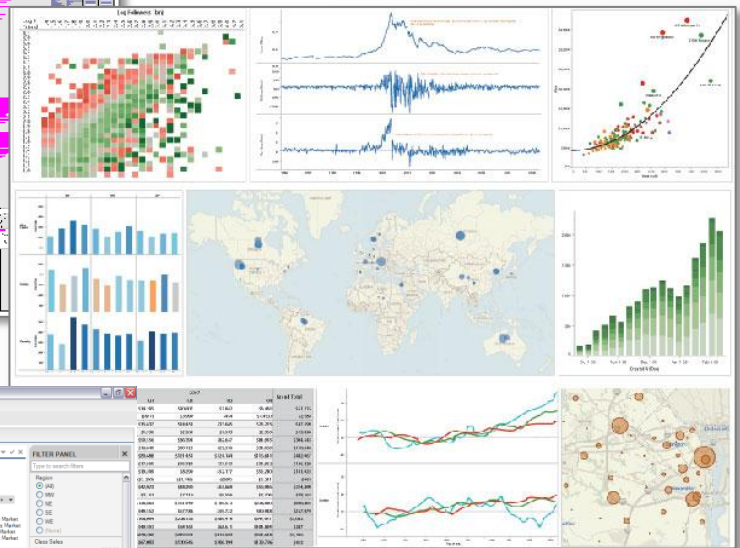
# Insight Discovery on Tabular Data: Example

- NSF Grants
  - Grant Title
  - Amount
  - Date
  - Program Manager
  - Awardee / Researcher Name
  - Awardee Affiliation
  - ...

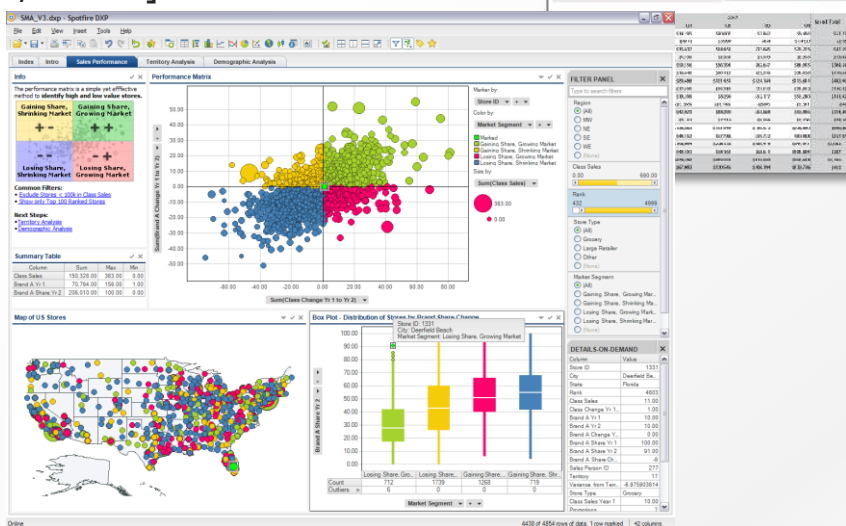
# Visualizing Tabular Data



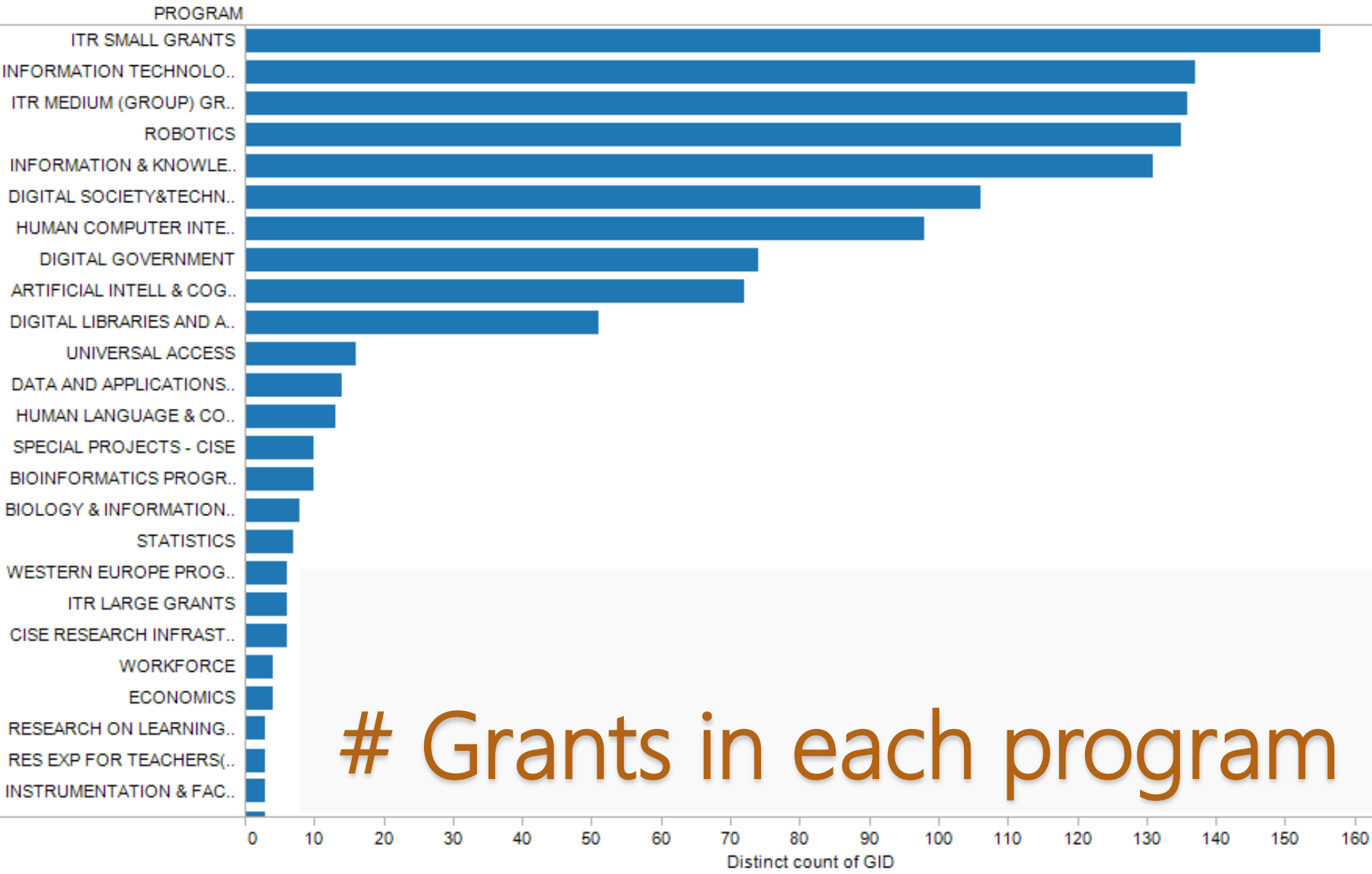
[Rao and Card, 1994]



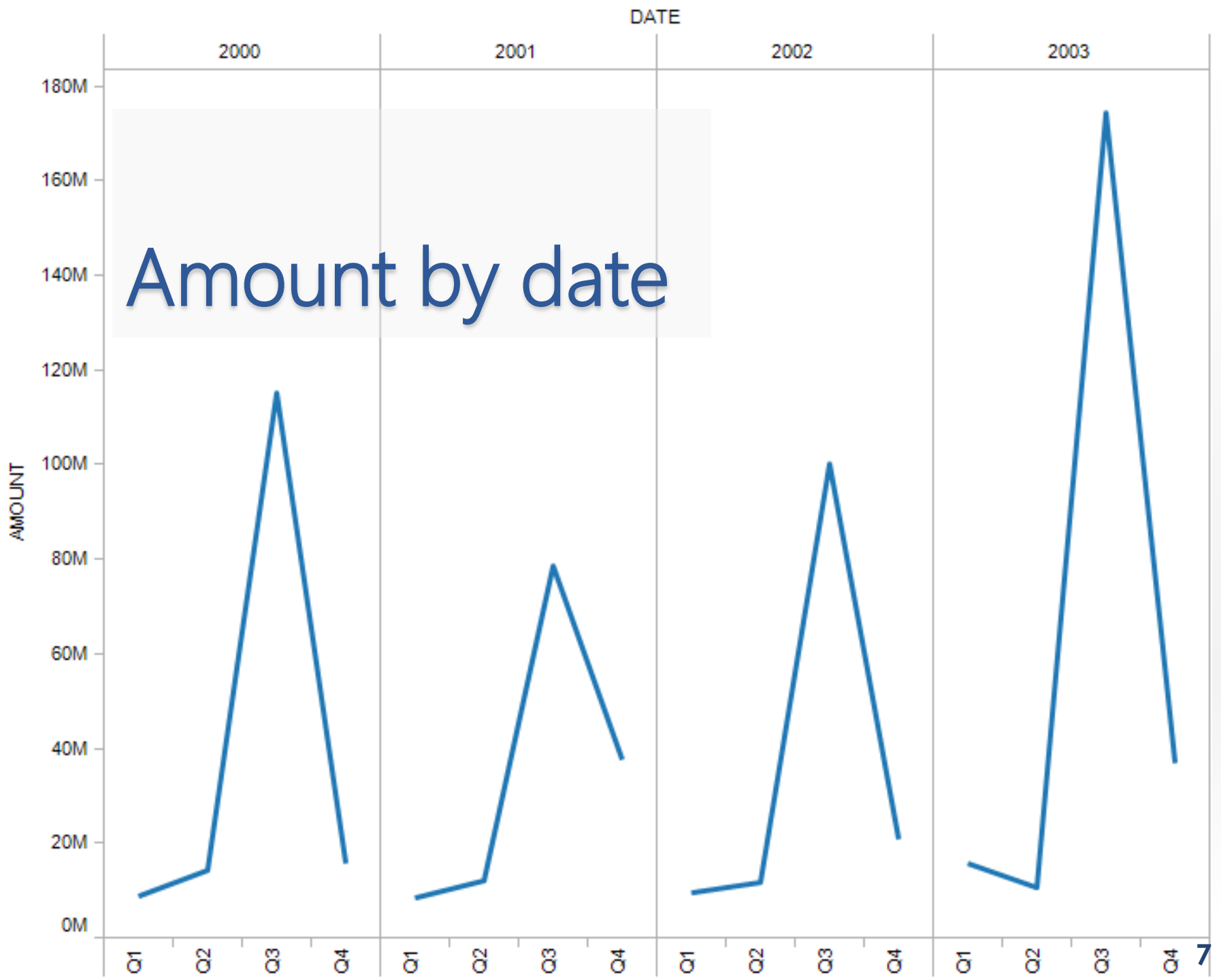
[Tableau]



[Spotfire]



# Grants in each program



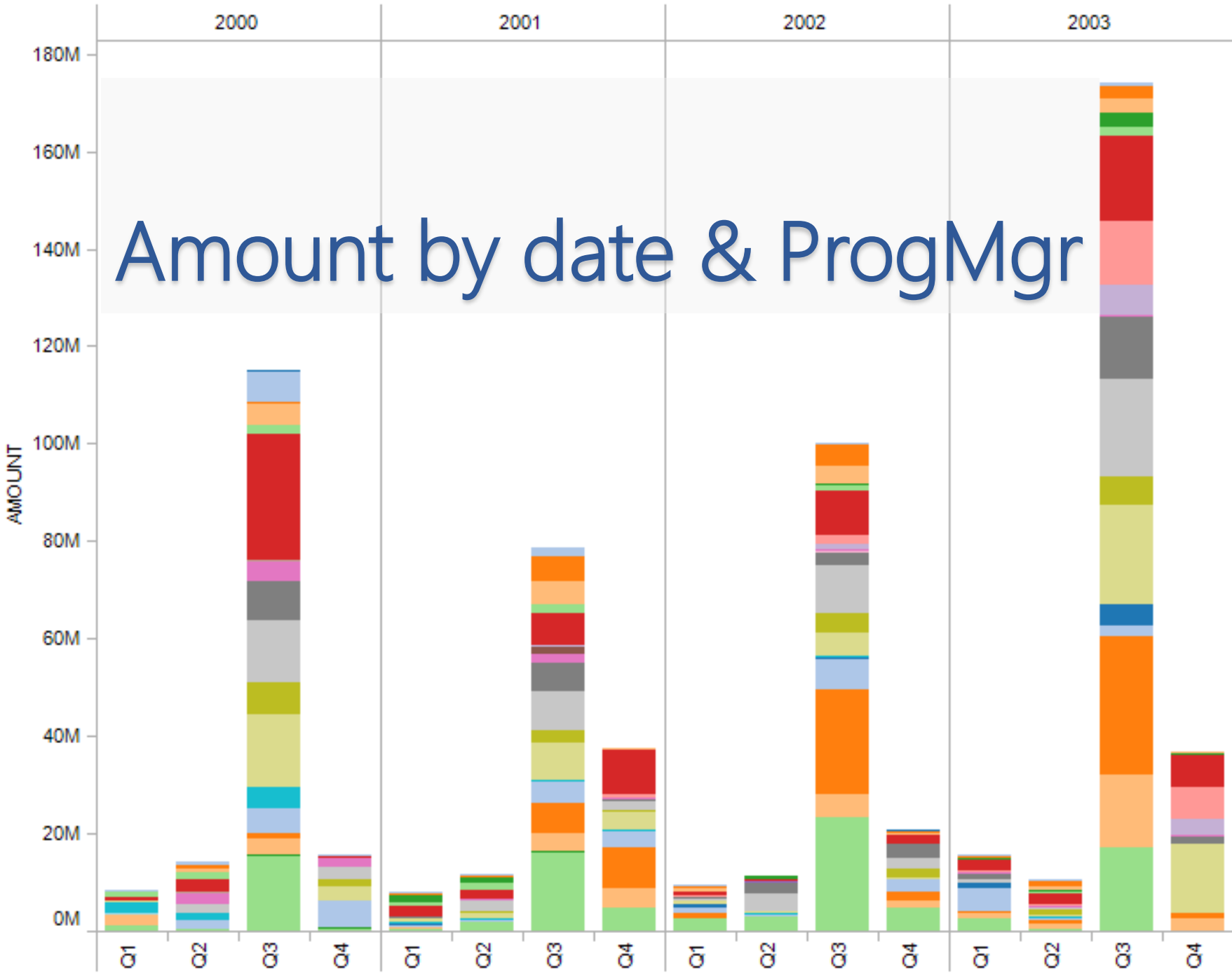
DATE

2000

2001

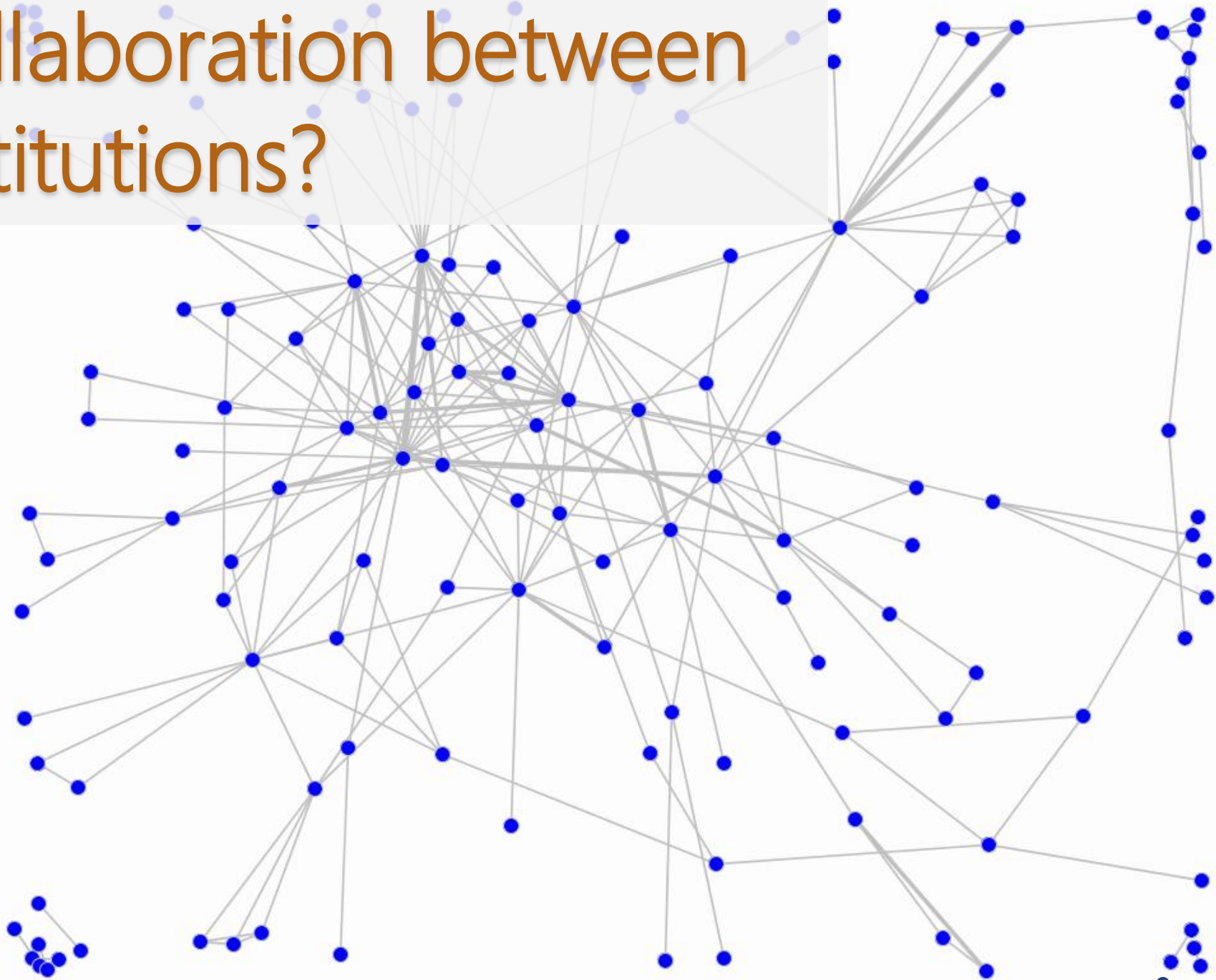
2002

2003

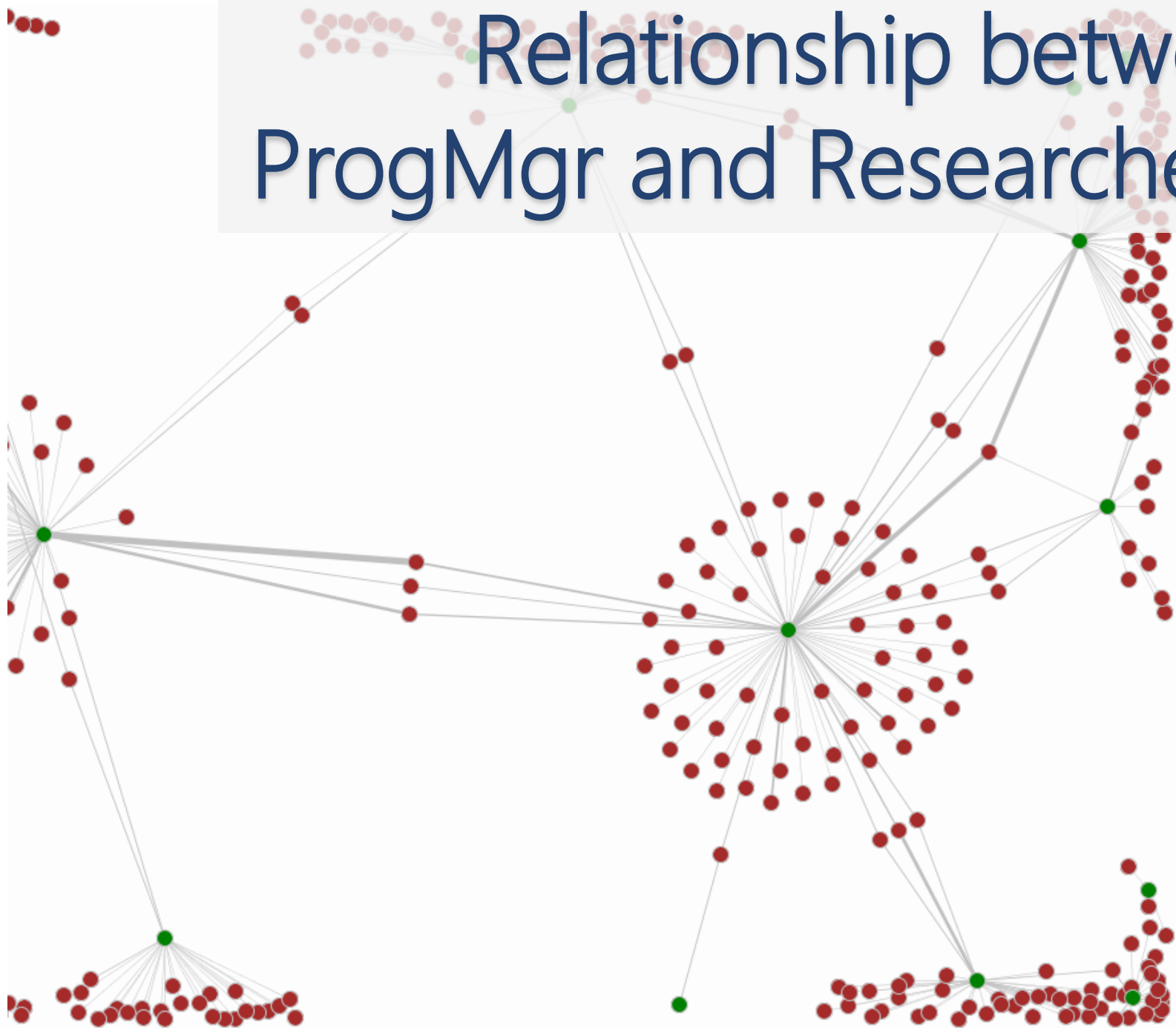




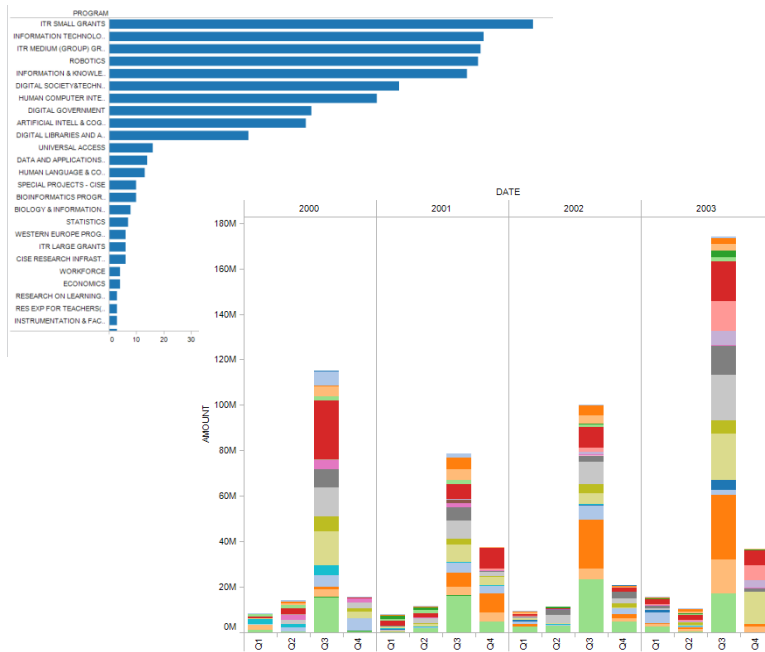
# Collaboration between institutions?



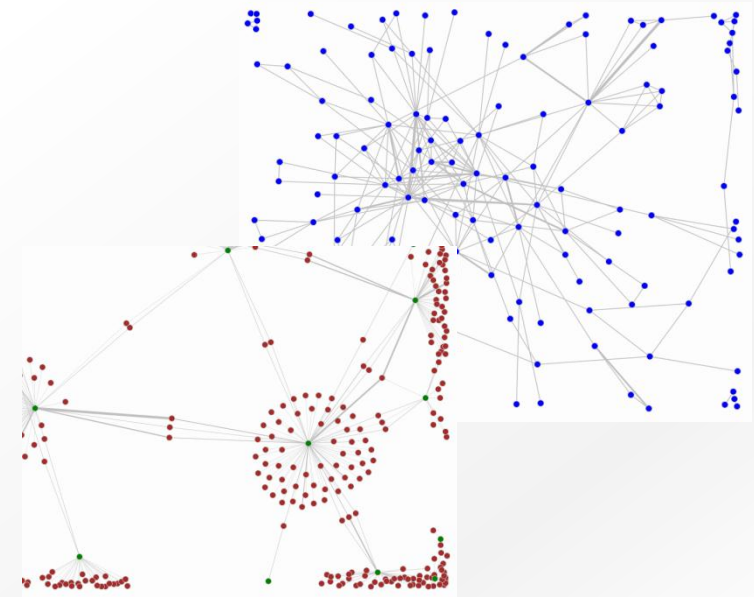
# Relationship between ProgMgr and Researchers?



- **Quantitative** attributes
  - Nominal data as independent variables, or not handled
- Patterns of distributions, correlations and outliers of numerical values



- **Nominal** attributes
  - Quantitative attributes are useful too
- Entities with interesting roles, emergent global structures

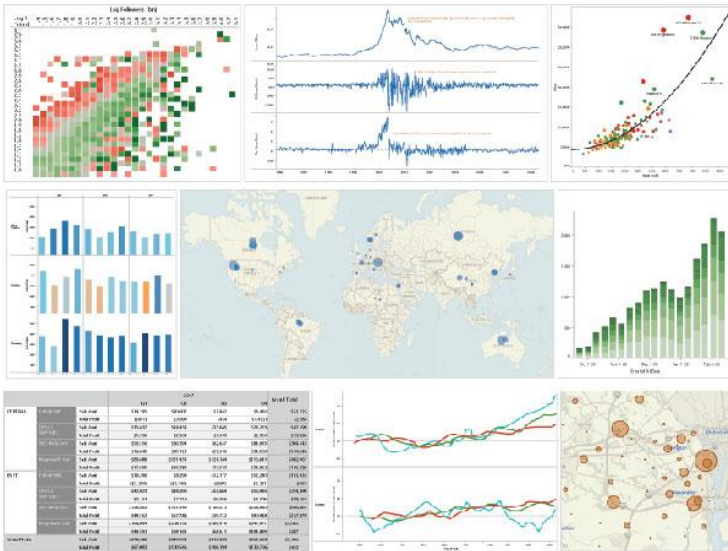


# Current State of the Art

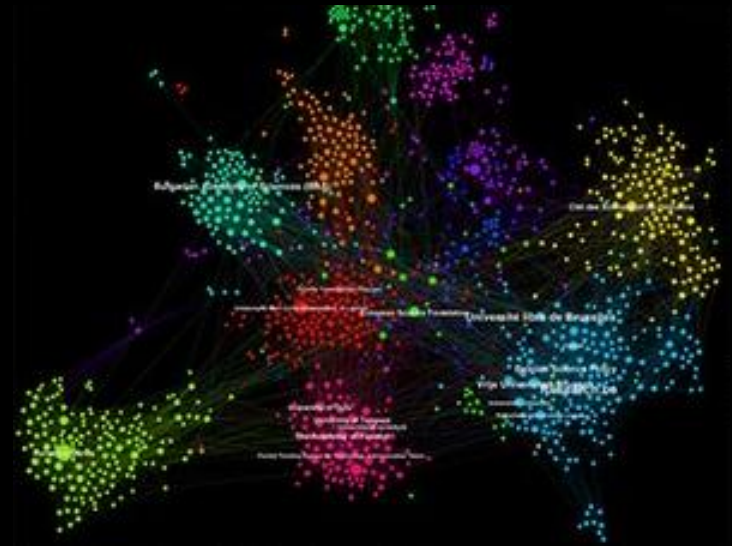
Tabular Data



Explicit Network



Spotfire , Tableau,  
TableLens ....



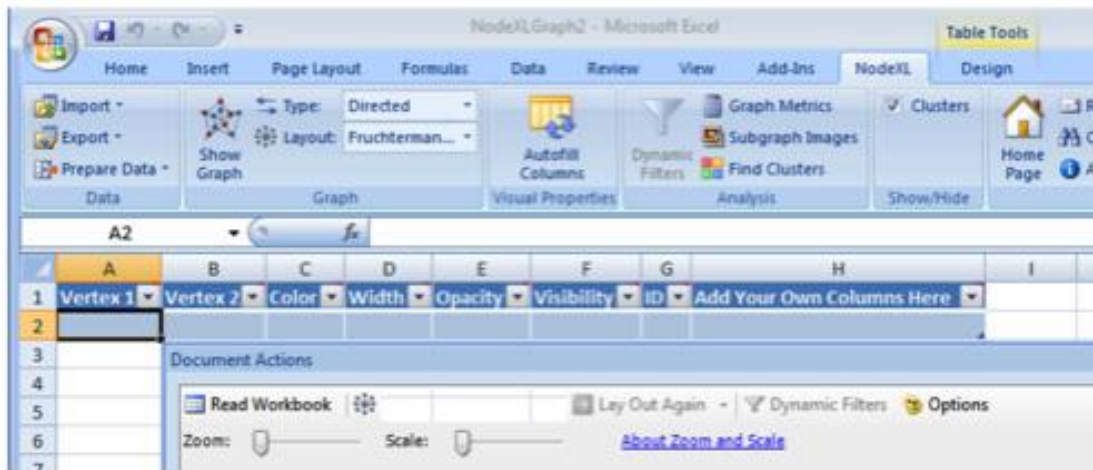
GUESS, UCINet,  
SocialAction ....

# Problems with this Partition

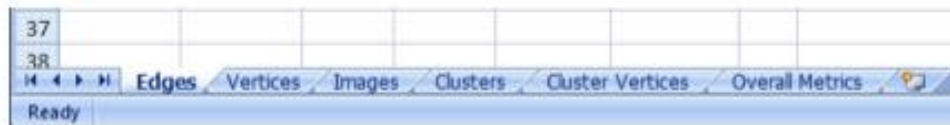
- Analytical: Network semantics are dynamic
- Usability: Modeling networks is tedious and requires programming skill

**Counter-intuitive for Exploratory Analysis**

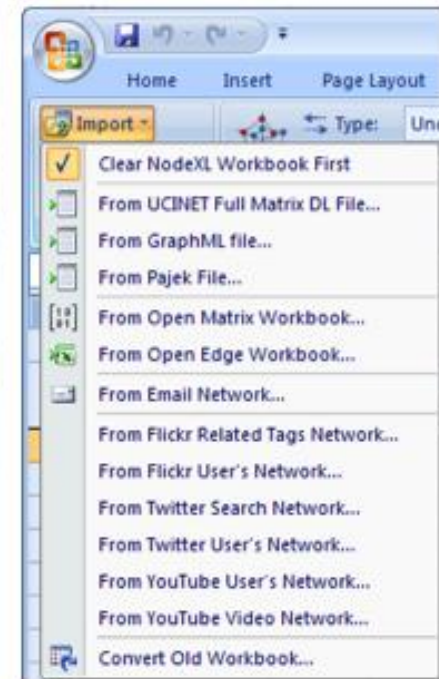
# NodeXL



(a) Specification of edge semantics



(b) Specification of network semantics



(c) Import Options

[Hansen, et. al. 2010]

# Questions & Approaches

## Goal: Network-based Visual Analysis of Tabular Data

1. Which *conceptually meaningful* operations are necessary to extract and transform tabular data into networks for exploratory analysis?

- Domain independent
- Generalized operations
- Expressive power

2. Given a set of operations, how to provide analysts with easy access to these operations and to couple network modeling with exploratory analysis?

- Hide technical details
- Reduce articulatory distance
- Immediate visual feedback

# Formal Framework


- Tables: Relational model [Codd, 1969]

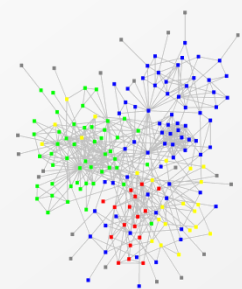
- Each row is uniquely identifiable
- Values in each cell is *atomic*: number, boolean, string, date

Product	Month	Region	Store	Target Sales	Actual Sales	Target Profit	Actual Profit	
Electronics	Dec-10	Atlanta	Bellevue	424100	395000	20000	22000	
Electronics	Dec-10	Atlanta	Condit	480000	278000	28800	9710	
Home Office	Dec-10	Atlanta	West	120000	146000	2000	14000	
Home Office	Dec-10	Atlanta	Tyus	80000	76000	2070	4630	
Home Office	Dec-10	Atlanta	Trent	420000	846000	3000	9500	
Industrial	Dec-10	Atlanta	Jay	48000	88000	3430	40	
Industrial	Dec-10	Atlanta	Alford	64000	102000	2730	3060	
Industrial	Dec-10	Atlanta	Frank	43000	76000	3040	3600	
Electronics	Dec-10	Atlanta	Jerry	53000	1000	2800	4970	
Electronics	Dec-10	Atlanta	Smith	48000	172000	1970	2100	
Electronics	Dec-10	Atlanta	Condit	428000	14000	3330	1730	
Home Office	Dec-10	Atlanta	West	120000	114000	2000	7400	
Home Office	Dec-10	All	EMPLOYEE_ID	FIRST_NAME	LAST_NAME	EMAIL	PHONE_NUMBER	HIRE_DATE
Industrial	Dec-10	All	100	Steven	King	SKING	515.124.4567	17-JAN-87
Industrial	Dec-10	All	101	Neena	Kochhar	NKOCHHAR	515.124.4566	21-SEP-89
Electronics	Dec-10	All	102	Lex	De Haan	LDEHAAN	515.124.4565	13-JAN-93
Home Office	Dec-10	All	103	Alexander	Hunold	AHUNOLD	590.423.4567	05-JAN-90
Industrial	Dec-10	All	104	Bruce	Ernst	BERNST	590.423.4566	21-MAY-81
Electronics	Dec-10	All	105	David	Austin	DAUSTIN	590.423.4569	26-JUN-97
Electronics	Dec-10	All	106	Valli	Pataballa	VPATABAL	590.423.4560	05-FEB-98
Electronics	Dec-10	All	107	Diana	Lorentz	DLORENTZ	590.423.5567	07-FEB-99
Home Office	Dec-10	All	108	Nancy	Greenberg	NGREENB	515.124.4568	17-AUG-94



- Networks: Weighted Simple Graphs

- Undirected 
- At most one edge between any two nodes
- Edges are weighted





# An Example

<b>ID</b>	<b>LastNm</b>	<b>FirstNm</b>	<b>Type</b>	<b>Date</b>	<b>Size</b>	<b>Visitee</b>	<b>Loc</b>
1	Dodd	Chris	VA	6/25/09	2018	POTUS	WH
2	Smith	John	VA	6/26/09	237	Office Visitors	WH
3	Smith	John	AL	6/26/09	144	Amanda Kepko	OEOB
4	Hirani	Amy	VA	6/30/09	184	Office Visitors	WH
5	Keehan	Carol	VA	6/30/09	8	Kristin Sheehy	WH
6	Keehan	Carol	VA	7/8/09	26	Daniella Leger	OEOB

# First-order Graph: Single Table

ID	LastNm	FirstNm	Type	Date	Size	Visitee	Loc
1	Dodd	Chris	VA	6/25/09	2018	POTUS	WH
2	Smith	John	VA	6/26/09	237	Office Visitors	WH
3	Smith	John	AL	6/26/09	144	Amanda Kepko	OEOB
4	Hirani	Amy	VA	6/30/09	184	Office Visitors	WH
5	Keehan	Carol	VA	6/30/09	8	Kristin Sheehy	WH
6	Keehan	Carol	VA	7/8/09	26	Daniella Leger	OEOB

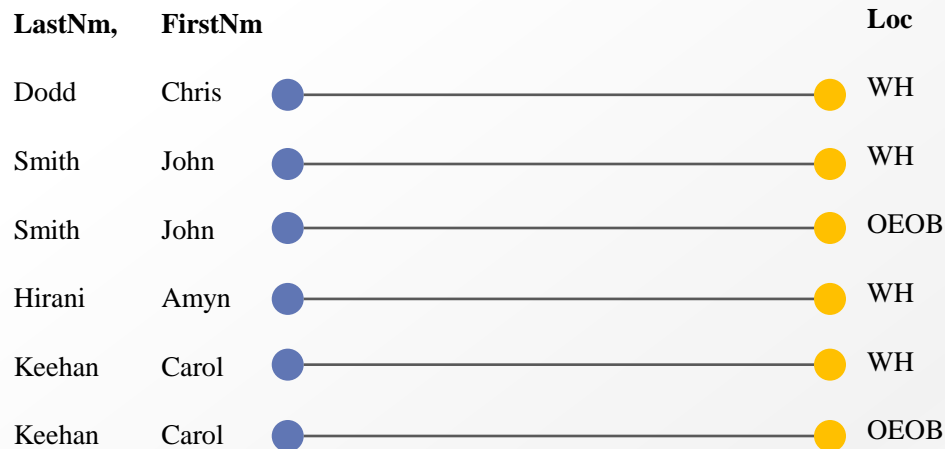
# First-order Graph: Single Table

ID	LastNm	FirstNm	Type	Date	Size	Visitee	Loc
1	Dodd	Chris	VA	6/25/09	2018	POTUS	WH
2	Smith	John	VA	6/26/09	237	Office Visitors	WH
3	Smith	John	AL	6/26/09	144	Amanda Kepko	OEOB
4	Hirani	Amy	VA	6/30/09	184	Office Visitors	WH
5	Keehan	Carol	VA	6/30/09	8	Kristin Sheehy	WH
6	Keehan	Carol	VA	7/8/09	26	Daniella Leger	OEOB

LastNm,	FirstNm		Loc
Dodd	Chris	●	● WH
Smith	John	●	● WH
Smith	John	●	● OEOB
Hirani	Amy	●	● WH
Keehan	Carol	●	● WH
Keehan	Carol	●	● OEOB

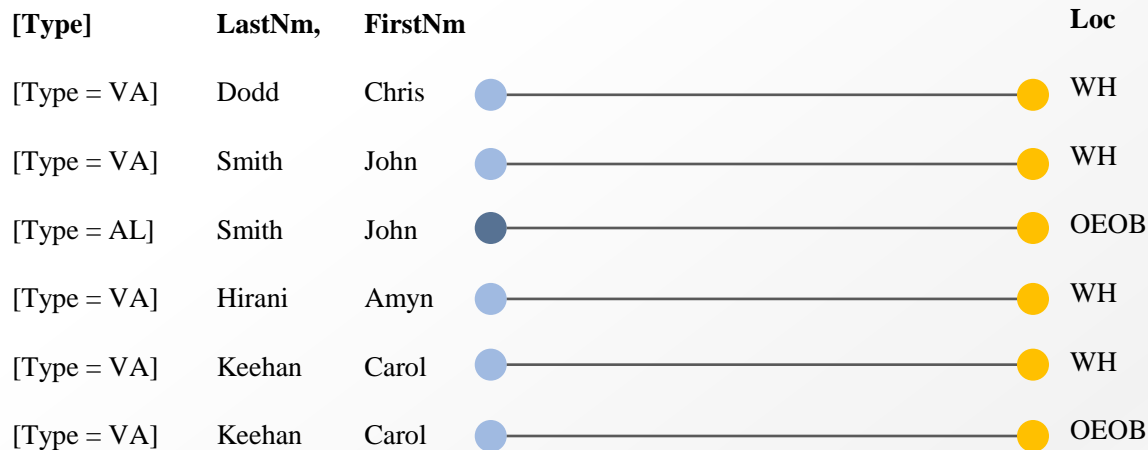
# First-order Graph: Single Table

ID	LastNm	FirstNm	Type	Date	Size	Visitee	Loc
1	Dodd	Chris	VA	6/25/09	2018	POTUS	WH
2	Smith	John	VA	6/26/09	237	Office Visitors	WH
3	Smith	John	AL	6/26/09	144	Amanda Kepko	OEOB
4	Hirani	Amy	VA	6/30/09	184	Office Visitors	WH
5	Keehan	Carol	VA	6/30/09	8	Kristin Sheehy	WH
6	Keehan	Carol	VA	7/8/09	26	Daniella Leger	OEOB



# First-order Graph: Single Table

ID	LastNm	FirstNm	Type	Date	Size	Visitee	Loc
1	Dodd	Chris	VA	6/25/09	2018	POTUS	WH
2	Smith	John	VA	6/26/09	237	Office Visitors	WH
3	Smith	John	AL	6/26/09	144	Amanda Kepko	OEOB
4	Hirani	Amy	VA	6/30/09	184	Office Visitors	WH
5	Keehan	Carol	VA	6/30/09	8	Kristin Sheehy	WH
6	Keehan	Carol	VA	7/8/09	26	Daniella Leger	OEOB

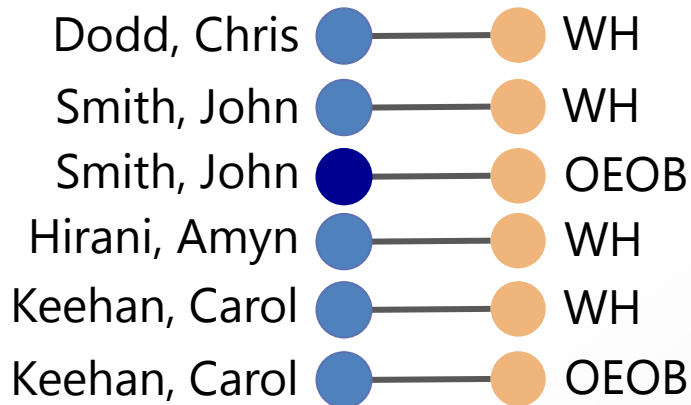


# Higher-order Graph: Transformations

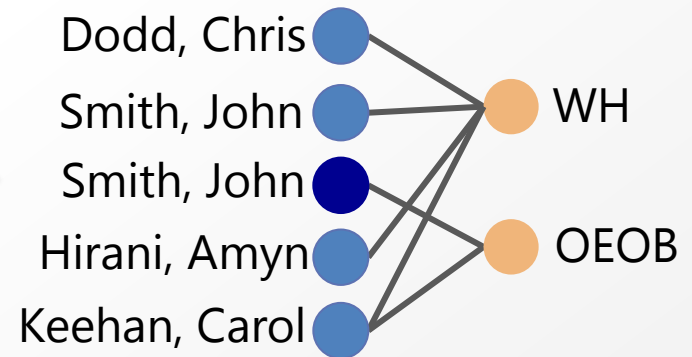
- Aggregation
- Projection
- Edge Weighting
- Slicing 'n Dicing

# Aggregation: Entity Resolution

**original graph**

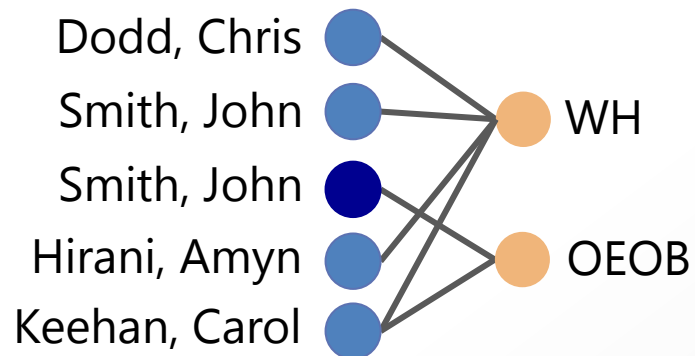


**after aggregation**

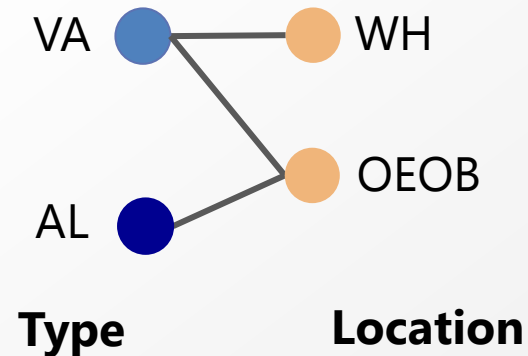


# Aggregation: Pivoting

**original graph**

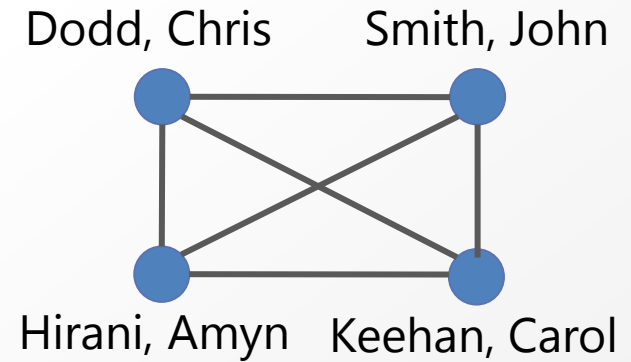
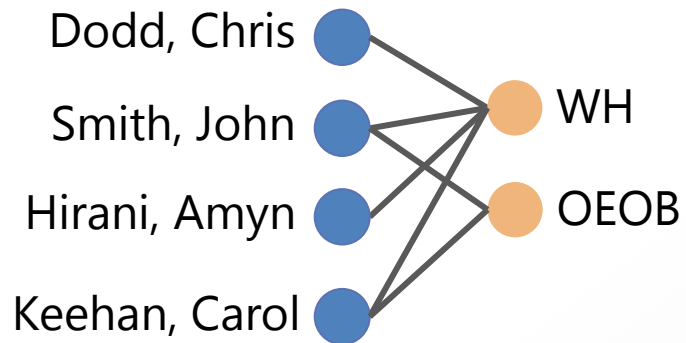


**after aggregation**



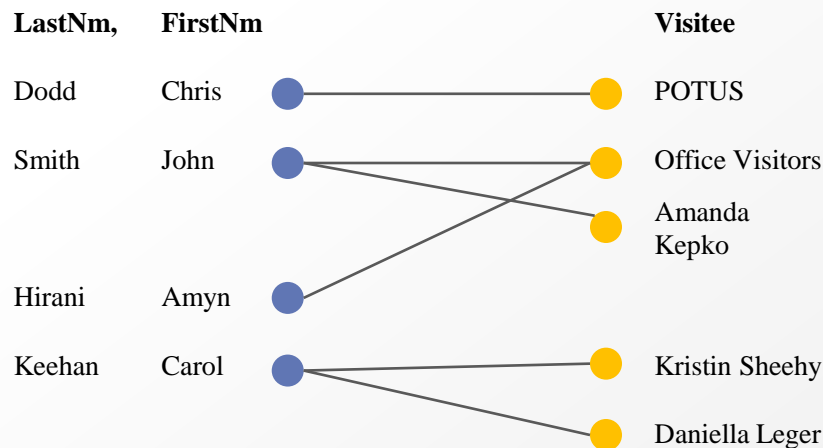


# Projection



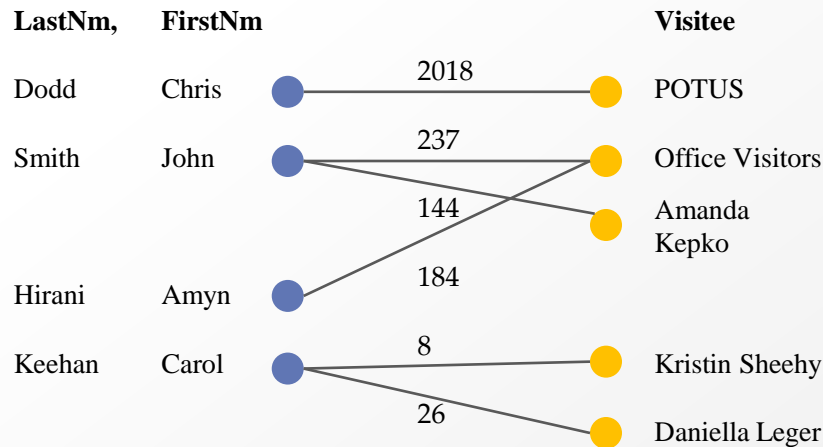
# Edge Weighting

ID	LastNm	FirstNm	Type	Date	Size	Visitee	Loc
1	Dodd	Chris	VA	6/25/09	2018	POTUS	WH
2	Smith	John	VA	6/26/09	237	Office Visitors	WH
3	Smith	John	AL	6/26/09	144	Amanda Kepko	OEOB
4	Hirani	Amy	VA	6/30/09	184	Office Visitors	WH
5	Keehan	Carol	VA	6/30/09	8	Kristin Sheehy	WH
6	Keehan	Carol	VA	7/8/09	26	Daniella Leger	OEOB



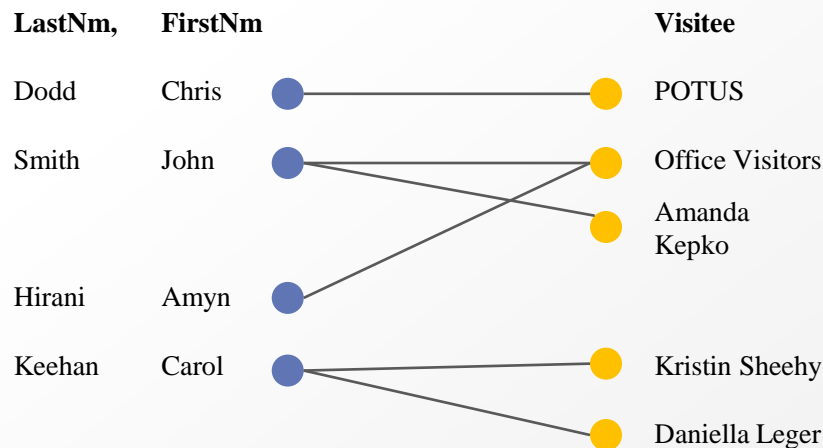
# Edge Weighting

ID	LastNm	FirstNm	Type	Date	Size	Visitee	Loc
1	Dodd	Chris	VA	6/25/09	2018	POTUS	WH
2	Smith	John	VA	6/26/09	237	Office Visitors	WH
3	Smith	John	AL	6/26/09	144	Amanda Kepko	OEOB
4	Hirani	Amy	VA	6/30/09	184	Office Visitors	WH
5	Keehan	Carol	VA	6/30/09	8	Kristin Sheehy	WH
6	Keehan	Carol	VA	7/8/09	26	Daniella Leger	OEOB



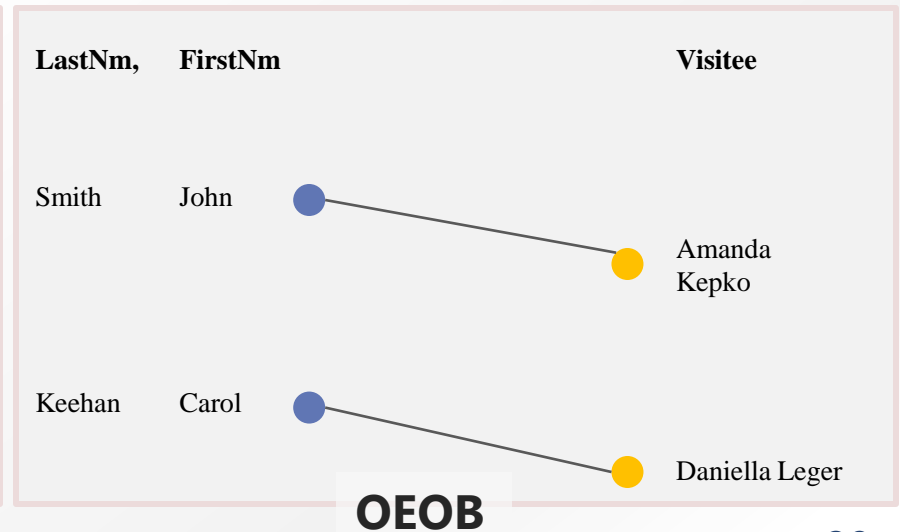
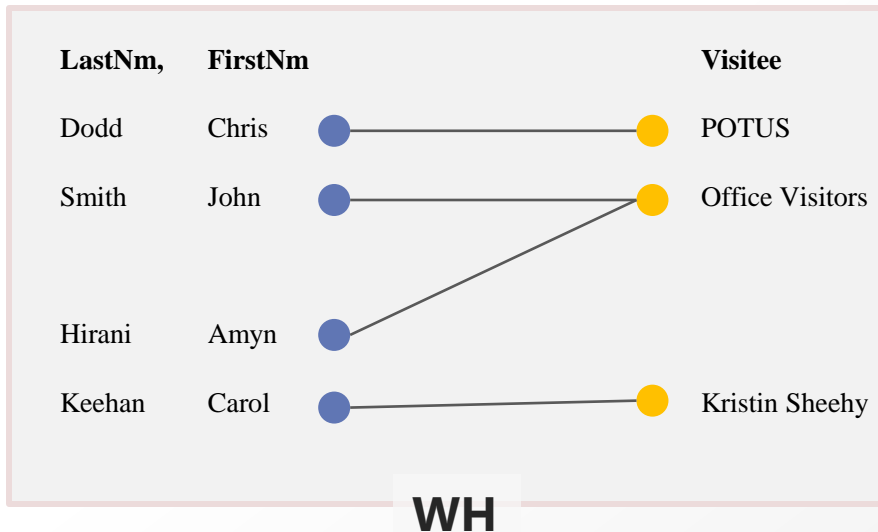
# Slice 'n Dice

ID	LastNm	FirstNm	Type	Date	Size	Visitee	Loc
1	Dodd	Chris	VA	6/25/09	2018	POTUS	WH
2	Smith	John	VA	6/26/09	237	Office Visitors	WH
3	Smith	John	AL	6/26/09	144	Amanda Kepko	OEOB
4	Hirani	Amy	VA	6/30/09	184	Office Visitors	WH
5	Keehan	Carol	VA	6/30/09	8	Kristin Sheehy	WH
6	Keehan	Carol	VA	7/8/09	26	Daniella Leger	OEOB



# Slice 'n Dice

ID	LastNm	FirstNm	Type	Date	Size	Visitee	Loc
1	Dodd	Chris	VA	6/25/09	2018	POTUS	WH
2	Smith	John	VA	6/26/09	237	Office Visitors	WH
3	Smith	John	AL	6/26/09	144	Amanda Kepko	OEOB
4	Hirani	Amy	VA	6/30/09	184	Office Visitors	WH
5	Keehan	Carol	VA	6/30/09	8	Kristin Sheehy	WH
6	Keehan	Carol	VA	7/8/09	26	Daniella Leger	OEOB



# Expressive Power

- Proximity grouping
- Extending to directed one-mode network
- Limitations

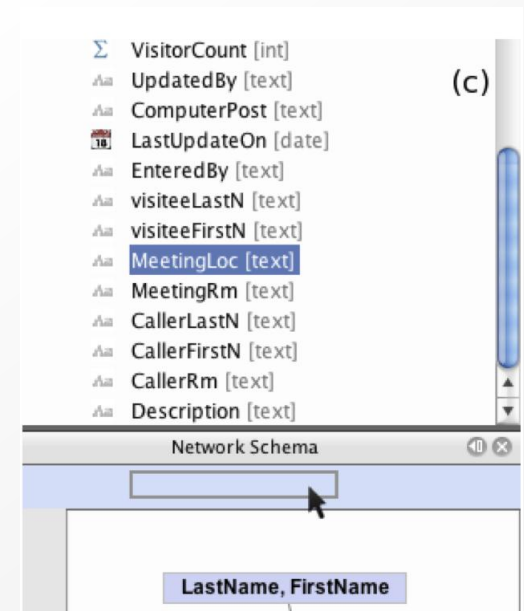
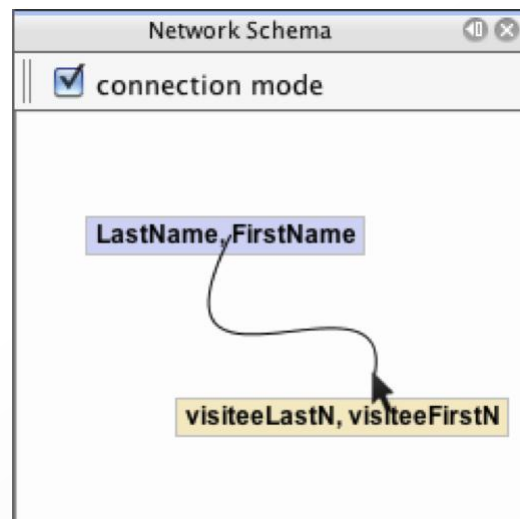
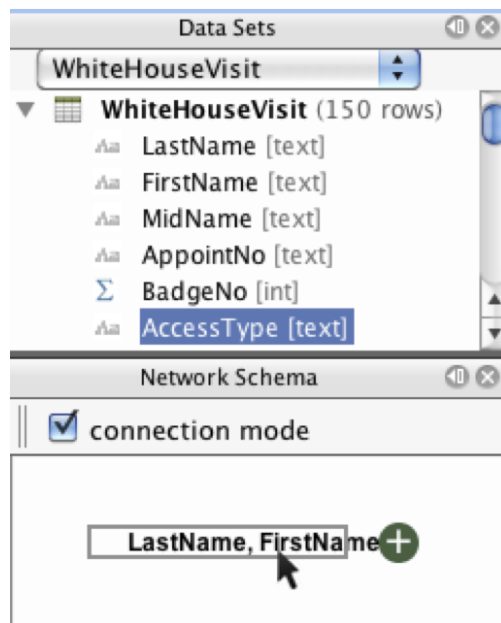
# Ploceus Interface Overview

The screenshot displays the Ploceus interface with three main views:

- Data Management View:** Shows a list of data sets: **Grants** (1264 rows) with fields `GID` [int], `title` [text], `date` [date], `program` [text], and `division` [text]; **Person** (1671 rows) with fields `PID` [int], `name` [text], and `organization` [text]; and **Works\_On** (2247 rows) with field `PID` [int].
- Network Visualization View:** Displays a network graph with 281 nodes and 862 edges. The layout is Fruchterman-Reingold. A search for "georgia" has been performed, resulting in 4 nodes highlighted in blue. A tooltip for these nodes lists: University of Georgia Research Foundation, Georgia State University, University of Georgia Research Foundation Inc, and Georgia State University Research Foundation, Inc. Three nodes are also highlighted with red boxes and labeled: University of Georgia Research Foundation, Georgia State University Research Foundation, Inc., and Georgia State University.
- Network Schema View:** Shows a simple graph with two nodes: `progmgr` and `organization`, connected by an edge.

At the bottom left, the "connection mode" checkbox is checked.

# Direct Manipulation Interface

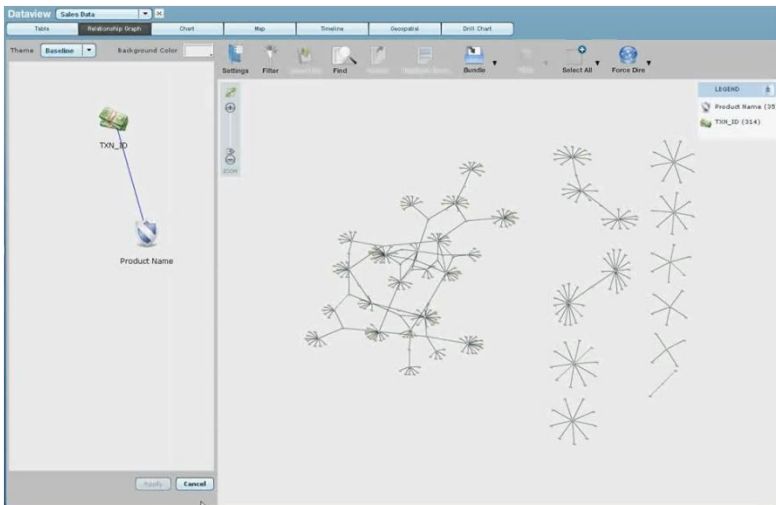




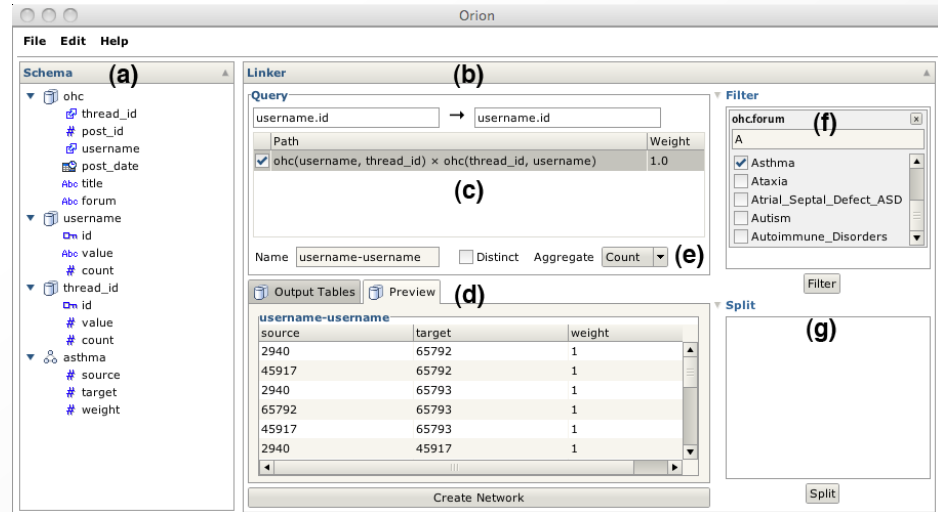
# Ploceus

Demo

# Related Work



**Centrifuge**



**Orion**

# Multiple Tables

<b>GID</b>	<b>Title</b>	<b>Program</b>	<b>Program Manager</b>	<b>Amount</b>	<b>Year</b>
1	Data Mining of Digital Behavior	Statistics	Sylvia Spengler	2241750	2001
2	Real-time Capture, Management and Reconstruction of Spatio-Temporal Events	Information Technology Research	Maria Zemankova	430000	2000
3	Statistical Data Mining of Time-Dependent Data with Applications in Geoscience and Biology	ITR for National Priorities	Sylvia Spengler	566644	2003

<b>PID</b>	<b>Name</b>	<b>Org</b>
1	Padhraic Smyth	University of California Irvine
2	Sharad Mehrotra	University of California Irvine

<b>Person</b>	<b>Grant</b>	<b>Role</b>
1	1	PI
2	1	coPI
2	2	PI
1	3	PI

# First-order Graph: Multiple Tables

GID	Title	Program	ProMgr	Amount	Year
1	Data Mining of Digital Behavior	Statistics	Sylvia Spengler	2241750	2001
2	Real-time Capture, Management ...	Information Technology Research	Maria Zemankova	430000	2000
3	Statistical Data Mining of Time-Dependent ...	ITR for National Priorities	Sylvia Spengler	566644	2003

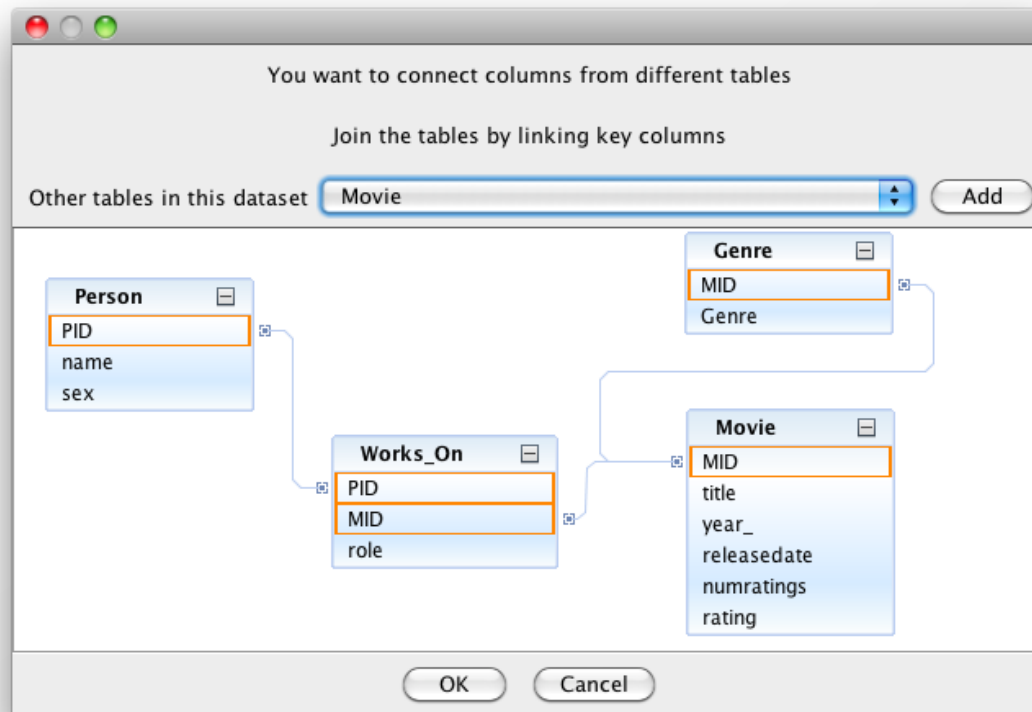
PID	Name	Org
1	Padhraic Smyth	University of California Irvine
2	Sharad Mehrotra	University of California Irvine

Grant	Role	Person
1	PI	1
1	coPI	2
2	PI	2
3	PI	1

Title	Program	ProMgr	Amount	Year	Grant	Role	Person	Name	Org
					1	PI	1		
					1	coPI	2		
					2	PI	2		
					3	PI	1		

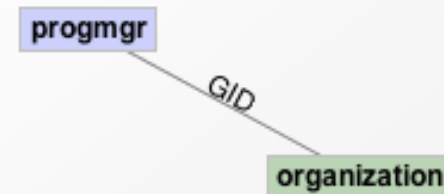
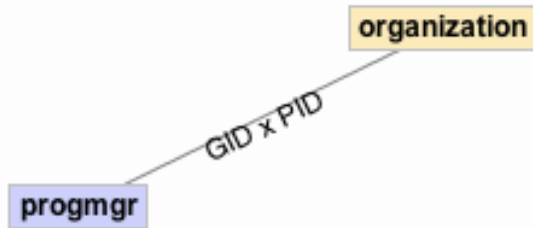
# Open Issues with Multiple Tables (1)

- Join Specification



# Open Issues with Multiple Tables (2)

- Interpretation of Edge Weights



# Acknowledgments



IIS-0915788



VACCINE Center