

Memo:

To: UIX Analysts & Future Gephi-users

From: Jessica McInchak

Date: August 2013

Re: Methods for data analysis and visualization using Gephi

TABLE OF CONTENTS 1. What is Gephi? 2. Preparing data a. Nodes and edges .csv tables 3. Importing data 4. Visualizing a. Basics b. Frequently used icons c. Ranking d. Layout e. Filters and queries f. Partition g. Statistics h. Preview i. Export 5. Sample visuals and network maps from UIX

1. What is Gephi?

- a. Gephi is an open-source platform to process, visualize, and perform statistical analysis on network data. The latest version can be downloaded for free here: <u>https://gephi.org/</u>
- b. This methodology uses the working example of the Urban Innovation Exchange (UIX) survey data collection. By August 2013, 82 innovators and their corresponding projects were surveyed and asked to list other organizations or people that were critical collaborators, community partners, funders, or the like. The objective of this data analysis is to capture the growing network of social innovation and entrepreneurship in Detroit and better understand intrapersonal support systems. (Ask Kat or Jessica for further context about this project if needed!)

2. Preparing data

- a. Most simply, network data must include "nodes" (in this example, unique project names and the people associated with them) and "edges" (how projects are connected to each other).
- b. Clean data to form a matrix that must include a "Source" project and a "Target" project with unique identifiers. (For UIX, this reads as "Fresh Media Group named Allied Media Project as a collaborator.)"

X 🖌	UK_projects_matrix_UUY - Microsoft Excel								
File	e Home Insert Page Layout Formulas Data Review View								
	🔏 Cut	Calibri - 10 - A A A = = = Wrap Text General	1	- 🧾 🧊 Normal	Bad	Good Neutr			
Paste	Format Painter	B I Ⅱ * □ * □ Δ * ▲ * ■言言 注 伊 圆 Merge & Center * \$ *	% , 50 .	Conditional Format Formatting * as Table *	Explanatory	Hyperlink Input			
c	lipboard (Font G Alignment G I	fumber	a	St	tyles			
	8571	▼ (* J* Woodbridge Pub		1					
	Α	В	С			D			
1	source	proj_name	target	relation_proj					
2	1	Fresh Media Group	47	Allied Media Project	:				
3	1	Fresh Media Group	108	B Detroit Design Festival					
4	1	Fresh Media Group	199	Motorcity Pride					
5	1	Fresh Media Group	18	Ruth Ellis Center					
6	1	Fresh Media Group	261	Wink					
7	3	Alternatives for Girls	84	Common Ground Sa	nctuary				
8	3	Alternatives for Girls	138	Enchanted Makeove	rs				
9	3	Alternatives for Girls	18	Ruth Ellis Center					
10	3	Alternatives for Girls	228	Starfish Family Serv	ces				

- c. Prepare two .csv files for Gephi: 1) nodes table; and 2) edges table.
 - A nodes table is most basically structured to have a unique Id and Label. Other attributes corresponding to each node can be included in columns to the right and later used as filtering characteristics. (In our example, this includes a binary flag field for UIX projects versus other un-related organizations named as community partners).
 - An edges table includes the Source and Target Id columns and edge "Type". Type can be "directed" (an arrow) or "undirected" (like reciprocal relationships). An additional "Label" column can be added to include edge names (for UIX, these were coded to distinguish advisor or mentorship from board or committee members and so on.)

	2 • D • La	(Z))	9-1-1-	_	UKBR	edges JULY Meruso	t Excel		Criffing And
File	Home Inset Ratelayout Farmulas Data Review View	28	Hore Steel	n Page Leyout	Formulas Data	Review View			a 😧 = 🖓 🗄
Paste	Cohn • II • A [*] A [*] = =	Peder	A Callee B J U and C	- 11 - A - 1 - 22 - funt	x = = = ⇒- Δ - = = = = ⇒- Alignment	ter Bill - 1/4 - 1/3 Number	Conditional Formations Conditional Formations Conditional Table * Conditional Condition Conditional Condition Conditional Condition Conditional Condition	P ^{al} Delete + P ^{al} Delete + Cem	2. Sont & Find & 2. Fitter - Select - coming
Clipbo	rd is Font is Alignment is Number is Styles Cells Editing		17 .	- fe	1 225				
	A1 • 🖉 🏂 Id	¥	A	В	C	D	E	F	G
	A B C	- 1	Source	Target	Туре				
1	d Label UIX	= 2	1	47	Directed				
2	1 Fresh Media Group 1	3	1	108	Directed				
3	2 Detroit Soup 1			100	Directed				
4	3 Alternatives for Girls 1	4	1	199	Directed				
5	4 Villages Community Development Corporation 1	5	1	18	Directed				
6	5 Pay It Forward Initiative 1	6	1	261	Directed				
7	6 Sunday Dinner Company 1	7	3	8 84	Directed				
8	7 Thrive Detroit 1	8	3	138	Directed		-		
9	8 Young Nation 1			100	Directed				
10	9 City Wings 1	9	3	18	Directed				
11	10 Mt Elliot Maker Space 1	10) З	3 228	Directed				
12	11 FoodLab Detroit 1	11	3	243	Directed				
13	12 Young Adults of Heidelberg 1	17	3	254	Directed				
14	13 HUGH 1	13	4	131	Directed				

3. Importing data (.csv) to Gephi

- a. Open Gephi and select "New Project" in the Welcome window
- b. Click the "Data Laboratory" tab \rightarrow "Data Table" \rightarrow "Import Spreadsheet"

Import spreadsheet	
Steps	General options
General options Import settings	Choose a CSV file to import: Separator: Comma Preview: Previ
	< Back Next > Einish Cancel Help

- c. Import both Nodes and Edges files; order of import does not matter. When choosing Nodes file, make sure that "As Table:" dropdown matches "Nodes Table." Repeat with Edges file.
- d. Continue through with default Import Settings.

e. When both files are imported successfully, toggle between "Nodes" and "Edges" under "Data Table" to display data. Save workspace at this time.

4. Visualizing

- a. Basics
 - 1) Switch from "Data Laboratory" tab to "Overview" tab. Imported raw data will display.



- 2) Find the number of displayed nodes and edges in the top right corner under "Context". The number of nodes should be the number of unique ids and labels imported. Edges should also be equivalent or may be smaller if there are multiple ties between the same source and target; rather, these are displayed as the thickest arrows.
- 3) Zoom using mouse scroll ball. Right click to adjust image in screen.
- b. Frequently used icons



c. Ranking

1) Adjust the size and color of nodes and edges based on "degree", meaning number of ties and connectedness.

	🌀 Gephi 0.8.2 - Project 0		
	File Workspace View Tools Wind	ow Plugins <u>H</u> elp	
	Overview	Data Laboratory	
	Ranking %		- Color: Size/Weight: Label Color: Label Size/Weight
Rank nodes or edges or both \rightarrow		<u> </u>	Color, Size, Weight, Eaber Color, Eaber Size, Weight Color, Size, Weight
	Degree	▼	
	Color:	■ ← Click	See preset color ramps
		Click on triangl	es to manually a li
	Range: U	Range of how	many ti-
	Spline	a Apply	E Apple
			APPLYI CHICHWORK
Show a color-coded			
results table \rightarrow	II Y		
d Lavout			
	Change the layour	t of the network m	ap This particular function takes a bit of
• • • • • • • • • • • • • • • • • • • •	experimenting and	tinkerina, especia	ally with big networks where nodes often overlap.
	The Gephi "Lavout	Tutorial" found he	ere is particularly helpful: https://aephi.org/users/
	Layout %		4
	Choose a layout	*	\leftarrow Choose from multiple layouts in the dropdown
	Choose a layout		menu.
	Clockwise Rotate	E	
	Concentric Layout Contraction		Note that not all options are a layout – some are
	Counter-Clockwise Rotate		adjustments to your current layout, like "Rotate."
	Expansion	-	
	Layout %		
	Force Atlas	•	← Select a lavout
Layout description \rightarrow	. 1	🕞 Run	← RUN! (And Stop your layout)
	Force Atlas		
	Inertia Repulsion strength	0.1 200.0	
	Attraction strength	10.0	
	Maximum displacement Auto stabilize function	10.0	Click on a category to see a short description at
	Autostab Strength	80.0	the bottom and adjust numbers to alter layout
	Autostab sensibility Gravity	30.0	display. Reset to defaults at bottom.
	Attraction Distrib.		
	Adjust by Sizes Speed	1.0	
	Penulsion strength		
	How strongly does each node reject (others	
	Presets Reset		

- 2) Extra lay-out tips -- consider installing additional plug-ins!
- e. Filters and queries

1) Use the "Filter" tab and complimentary "Query" function to control the display size of your network, this is useful for creating an accessible, readable image. The "Filter" tab is located on the right panel.



- 2) Clicking "Filter" twice will un-do the selected filter. Or right-click on any selected filter to "Remove."
- 3) If you are running multiple filters and queries, it becomes helpful to export to a new workspace after each filter (the middle icon next to "Reset" below "Filters" tab). This also re-calculates the network rankings as they change through filters. For example, certain UIX projects have the greatest out-degree when analyzed in the full-network, but others are highest in the filtered network of only UIX-featured projects.

f. Partition

1) Partitioning happens in two places in the Gephi workspace; first, as a filter and then as a nodes and edges color/size parameter.



2) Once you set a query for an attribute using "Partition", you can also control "Partition" in your layout. Go to "Window" drop-down and choose "Partition." It should appear as a tab next to "Ranking" on the left.

Gephi 0.8.2 - uix_projs_edg	es 0808 gephi	
File Workspace View Tools	Window Plugins Help	
Overview	🔲 Data Laboratory 📃	
Partition 📽 Rai king		
Nodes Edges		ϵ From the "Partition" tab, you can toggle between Nodes
<i>©</i>	-	and Edges, depending on your selected attribute.
		Since we queried "Partition (Tie)" in our "Filter", we see "Tie" in the "Edges" dropdown. You may have to "Refresh" first.
		Partition % Ranking Nodes Edges Image: Choose a partition parameter Choose a partition parameter The
Group Group Show Pie	Apply	

Partition % Ran	nking		
Nodes Edges			
🤣 Tie			•
collab_tie		(33	2.63%)
collateral_tie		(11	1.82%)
fund_tie		(10).23%)
any_tie		(8.	64%)
support_tie		(6.	35%)
board_tie		(5.	29%)
space_tie		(5.	11%)
two_ties		(5.	11%) =
resource_tie		(4.	23%)
tech_tie		(4.	06%)
media_tie		(2.	12%)
market_tie		(1.	41%)
advisor_tie		(1.	41%)
legal_tie		(0.	88%)
design_tie		(0.	71%) 👻
G Show Pie			Apply

← Choose "Tie" to assign unique colors to network edges based on an attribute.

Right-click anywhere in the window to "Randomize" the default color spectrum.

OR right-click an individual color to set RGB.

 \leftarrow Use "Show Pie" to see a chart break-down of the selected edge attribute.

- g. Statistics
 - 1) Beyond layouts, Gephi also performs high-speed, advanced social network analysis. Statistics about the network's connectivity can be calculated and exported as a .csv.

Statistics % Filt	ters				-							
Settings												
Network Over	rview											
Average Degree			1.116	Run	3							
Avg. Weighted De	gree			Run	۲	\leftarrow Run a calculati	ion. See	an ir	nstar	nt "Rep	ort."	Find
Network Diameter			Avg. Weight	ed De	gree	newly stored date	a in "Dat	a Lal	bora	tory" →	• "Da	ıta
Graph Density				Run	۲	Tables" \rightarrow "Nodes	s."					
HITS				Run	۲							
Modularity				Run	۲	Run numerous ca	llculation	ns as	desir	ed, the	y will	build
PageRank				Run	۲	more columns to	ine ien ii	n ine	INO	des ic	ible.	
Connected Compo	nents			Run	۲							
☑ Node Overvie	w											
Avg. Clustering Co	efficient			Run	۲							
Eigenvector Centra	ality			Run	۲						aree	3
💌 Edge Overvie	w									"D	SQL	a."
Avg. Path Length				Run	۲			, colu	mns`	with "Export	TODI	<u>,</u>
						K NC	otice new Istics and	optic	on is			
Data Table 🕫 Graph	88					ςιo					ſ	< > •
odes Edges 🛛 🖓 Configura	ation 🚯 Ad	id node H Add edg	e 🃸 Search/Replace	📳 Imp	ort Spre	adshee 🖳 Export table 🙀 More actions 🗸		Filter:		Nodes		- 9
Nodes				Id		Laud			UIX	In-Deg Out	-De Degr	ee
Fresh Media Group				1		Fresh Media Group			1	0	5	5 🔺
Detroit Soup				2		Detroit Soup			1	1	0	1
 Alternatives for Girls 				3		Alternatives for Girls			1	2	6	8
Villages Community Devel	lopment Corpo	ration		4		Villages Community Development Corporation			1	3	7	10
 Pay It Forward Initiative 				5		Pay It Forward Initiative			1	0	7	7
Sunday Dinner Company				6		Sunday Dinner Company			1	0	6	6
Thrive Detroit				7		Thrive Detroit			1	0	9	9
Young Nation				0		Veues Nation				0	0	-

- 2) Quick definitions of common metrics for statistical analysis:
 - 1. Centrality most fundamental term and used in sociology to indicate how well a node is connected
 - 2. In-degree how many edges point to a node or how "inwardly" connected a node is
 - 3. Out-degree how many connections a node points to or a node's "outward" connections
 - 4. Degree in-degree plus out-degree
 - 5. Modularity community detection for hierarchal data
 - 6. Reciprocity reciprocal ties indicate stability

- h. Preview
 - 1) Control the aesthetics of your visual output in "Preview" after setting the layout, filters, partitions, and/or ranks in "Overview." Preview offers better control for labeling and font sizes especially.

	the second s	-			
Overview	Data Laboratory		.	Preview	
Preview Settings 🚳		-	60 Previe	w ¥	
🕐 Presets		ы	0		
Default		•			
Nodes					
Border Width	1.0				
Border Color	custom [0,0,0]				
opacity	100.0				
Node Labels					
Show Labels					
Font	Arial 12 Plain				
Proportional size	3				
Color	custom [0,0,0]				
Shorten label					
Max characters	30				Use these controls to control visual network
Dutline size	0.0				
Dutline color	custom [255,255,255])			display. Defaults (important adjustments include
Dutine opacity	80.0				aispidy. Derduiis/impondni dajusimenis include
Box					autored and directed address addre aplan (set to
Box color	parent				curved of difected edges, edge color (set to
Box opacity	100.0	-			We wind the set of the
Edges					Tonginal II you have re-colored your edges
Show Edges			L		the second
Thickness	1.0		ſ		inrough a parimon), and hode labeling aisplay.
Rescale weight					
Color	mixed				
Opacity	100.0				
Curved					Click "Refresh" after any adjustments to see the
Radius	0.0				na sudd
Edge Arrows					result.
Size	3.0				
Edge Labels					
Show Labels	(m)				
Font	Arial 10 Plain				"Reset Zoom" at the bottom left of the display
Color	original	ŏ			
Shorten label					screen to re-center adjusted graphs.
Vax characters	30				
Dutine size	0.0				
Outline color	custom [255 255 255]				
Outline onacity	80.0				
Dutline opacity Preview ratio: 100%	80.0				

- i. Export
 - 1) Gephi exports images in three formats: .pdf, .png, and .svg. You can also export a spreadsheet from the Data Laboratory as a .csv with newly calculated statistics.
 - 2) Note that long labels or distant nodes may be cut off during the export process for images. To avoid this, choose "Options" in the export window and adjust page orientation, widen margin widths, or select transparency.



5. Sample visuals and network maps from UIX

- a. This is a 12-step guide to produce 6 images for the UIX network, illustrating the full network and filtering down to only UIX projects.
- b. The steps are illustrated with screen shots and the bounded pdf exports follow. All images were produced in the same work session.
- c. This series of networks is meant to show ways to use Gephi to think about "zooming in" on particular aspects of a network through filtering. The images also exemplify the two distinct levels of analysis in the UIX data (the entire network of 550+ organizations and the survey sample of 82 UIX-featured projects)

Step1: Import data. No Queries.

Step 2: Rank nodes by size and apply Circular Layout by UIX Attribute

		Gephi 0.8.2 - Project 0	
		<u>File Workspace View Tools Window Plugins Help</u>	
		Overview Data Laboratory	
J	Context 🕺	Partition Ranking %	-
]	Nodes: 536	Nodes Edges 🛛 🕥 💎 🆄	
	Edges: 601	Degree	
	Directed Graph	Degree	
		Min size: An Max size: The A	
	Statistics Filters %		
	Reset A+	Range:	
	Library	0 30	
	Attributes	Spline Apply	51
	tel		_
	Operator		
	🗈 🖷 Topology		
	Saved queries		
		■ ▼	
		Layout %	
		Circular Layout	-
	Q Drag filter here	i)	
		Node Placement	
		Node Layout Direction Counter Clockwise	-
		Prevent Node Overlap	- 11
		Excel Diameter	
		Diameter size 500.0	-11
		Order Nodes by (decreasing) UIX (Attribute)	-
		Enable Transition	$- \parallel$
		Tansiion Sieps 1000000	
		Circular Layout	
	Select Filter	Presets Reset	
	Workspace 0 🖪 🕨		

Step 3: Partition nodes by UIX binary, assign colors. Apply Noverlap layout after Circular Layout. Partition edges by Tie, assign colors. No queries are used yet. Once we assign colors and size, these setting will carry through to all future steps unless noted to "re-apply".

Overview Data tition % Ranking es Edges UIX 0 1	(84.7%) (15.3%)	Overview Partition % Ranking Nodes Edges	Data Laboratory
tition % Ranking es Edges UIX 0 1	(84.7%) (84.7%) (15.3%)	Partition # Ranking Nodes Edges Ø Tie	
es Edges UIX 0 1	(84.7%) (15.3%)	Nodes Edges	
UIX 0 1	(84.7%) (15.3%)	🛞 Tie	
0	(84.7%) (15.3%)		
0	(84.7%) (15.3%)	and the Ka	
1	(15.3%)	Collab_tie	(32.61
		collateral_tie	(11.65
		fund_tie	(10.65
		any_tie	(8.32%
		resource_tie	(6.66%
		support_tie	(6.49%
		space_te	(6.16%
		board_be	(5.329
		advisor tie	(4,237
		media tie	(2.339
		market_tie	(1.5%
		legal_tie	(1%)
		design_tie	(0.67%)
Group in Show Pie			
		Show Ple	
vorlan			
veriap	•		
	▶ Run		
loverlap			
3.0 3.0			
0 1.2			
gin 12.0			
verlap	0		
	-		

Step 4: Adjust aesthetic settings in Preview. **Export first image** of full network with color-coded edges and nodes.



Step 5: To go from the full network to the "zoomed-in" arc display, drag the Partition(UIX) to Queries and select the flag field of UIX = 1. Do not adjust layouts, rankings, or colors.



Step 6: Preview, reset zoom, and export image 2. Do not change any previous aesthetic settings.

File Workspace View To	ols <u>W</u> indow Plugins <u>H</u> elp		
Overview	Data Laboratory	Preview	
Preview Settings #		Già Preview 30	
📌 Presets		10 C	
Default			
Nodes			Ø Ø 0
Border Width	0.5		
Border Color	custom [255,255,255]		
opacity	100.0		
Node Labels			
Show Labels	1		
Font	Carson Christ & Fair		
Proportional size			
Color	custom [255,255,255]		
Shorten label	8		
Max characters	30		
Outline size	0.0		
Outline color	custom [255,255,255]		0 ^w •
Outline opacity	80.0		
Box	E1	0.00	
Box color	perent		
Bex opacity	100.0		() • ·
Edges			• •
Show Edges	(V)		© o
Thickness	1.5		
Rescale weight	(F1)		
Color	original		
Opacity	100.0		
Curved	E3		
Radius	0.0		
Edge Arrows			
Size	8.0		
Edge Labels			
Show Labels	123		
Font	Arial 10 Plain		
Color	original		
Shorten label			
Max characters	30		
Outline size	0.0		
Outline color	custom [255,255,255]		
Preview ratio: 100%	00 0	a)	
Carde Dic DOE DWC	o Kerre		
export: SVG/PDF/PNG		Background Reset zoom - +	
			Workspace 0 4 H

Step 7: Return to Overview and export the filtered graph to a new workspace. Once in a new workspace, re-apply Ranking of node size and color. Re-apply Noverlap layout only. This now displays the degree of connectedness between only UIX-featured projects.



Cashi 0.82

Step 8: Preview and reset zoom. **Export the third image**. As a pdf, this should appear just like the second image, besides node size and color. This way, they offer a comparison of two levels of analysis.

deprint d.Z - Project u	le contra c		
Elle Workspace View I	ools <u>W</u> indow Plugins <u>H</u> elp		
Overview	Data Laboratory	Preview	
Preview Settings #		Co Preview	4
Presets		141 I	36
Default		•	
Nodes			
Border width	0.5	0	
border Color	custom [255,235,255]	lend	
opacity	20010		
O NOGE LADER	1991	1.0	• • • • • • • • • • • • • • • • • • •
Show Labels	(M)	0	a de la
Font	and the state of the	(m)	
Proportional size	MI STATE AND AND		
Color	custom (255,255,255)	Les	9
Shorten soel			
Max characters	30		
Outrie SiZe	0.0	0	
Outline color	custom [255,255,255]	Lens	0
Outline opacity	30.0		
Box		0	
Box CORF	parent	(00)	
bex opacity	100,0		
C Edges	1941		
Show Edges	1		
Inickness	1.3		
Hescale weight	EJ		
Color	ongha		
Opacity	100.0		
Curved			
Radius	0.0		
Edge Arrows			
Dize	0.0		
Coge Labels	1000		
Show Labels		0	
Font	Anal to Plain		
Color Chartee Ishai	ingra	. Gest	
diversion and	10		
Duffine size	0.0		
Outline size	juli	0	
Outline court	en n	- Les	
Preview ratio: 100%	0		
	U 🛞 Re	efresh	
Export: SVG/PDF/PNG		Backgro	J Reset zoom - +

Step 9: With a new image of UIX-only projects sized by their connectedness to each other, now make the image more readable through a new layout. Re-apply Circular Layout, Nooverlap, and Rotate.

Circular Layout		•	
0		👂 Run	
Node Placement			
Node Layout Direction	Counter Clockwise	•	
Prevent Node Overlap			
Circle Properties			
Fixed Diameter			
Diameter size	500.0		
Order Nodes by (decreasing)	UIX (Attribute)	•	
Transition			
Enable Transition			
Fransition Steps	100000.0		
			0
		0	size
Circular Layout			A

Noverlap		•
0		🕞 Run
Noverlap		
speed	3.0	
ratio	1.2	
margin	12.0	
Noverlap		0

Step 10: Preview and Export image 4.



Step 11: Many nodes on this graph do not actually have any connections, but appear because they are UIX-featured projects. We can remove these by running a Topology –Degree Range query between 1-6.



Step 12: Preview and **export the fifth and final image** of only nodes with an edge connections! Now, to highlight each type of tie individually, export again to a new workspace and use the Partition (Tie) filter, and export each desired image (**Images 6+**).













