



Network Analysis of Ethereum Exchanges

Presented by

Team 4

**Soubhik Chakraborty, Monika Meshram, Lalitha Voruganti,
Shweta Dhakare & Prachi Vadhani**

Introduction - Blockchain

2

- ▶ Blockchain is a shared database that stores data in blocks connected through cryptography, differing from traditional databases in its architecture.
- ▶ While blockchains can store various types of data, the most prevalent use case is as a ledger for transactions.
- ▶ Decentralized blockchains are immutable, meaning that once data is entered, it cannot be altered or deleted. In the case of Bitcoin, transactions are permanently recorded and publicly visible.

Introduction – Anti-Money Laundering

- ▶ The main goal of Anti Money Laundering (AML) efforts is to prevent criminals from concealing the profits of their unlawful activities.
- ▶ Criminals use money laundering to make their illicit funds appear as though they were obtained legally.
- ▶ AML regulations mandate that financial institutions establish complex customer due diligence plans that assess money laundering risks and detect suspicious transactions.

EDA

Observations & Variables

Size – 8GB

One Instrument – 100,000 entries & 14000 edges

20 instruments – compressing the database to 22MB

Purpose of Our Analysis

- ▶ Classify groups, and segment data to search through millions of transactions to find patterns and detect fraud
- ▶ Learn suspicious-looking patterns and use those patterns to detect them further through neural networks
- ▶ Identify characteristics found in fraud through Pattern recognition

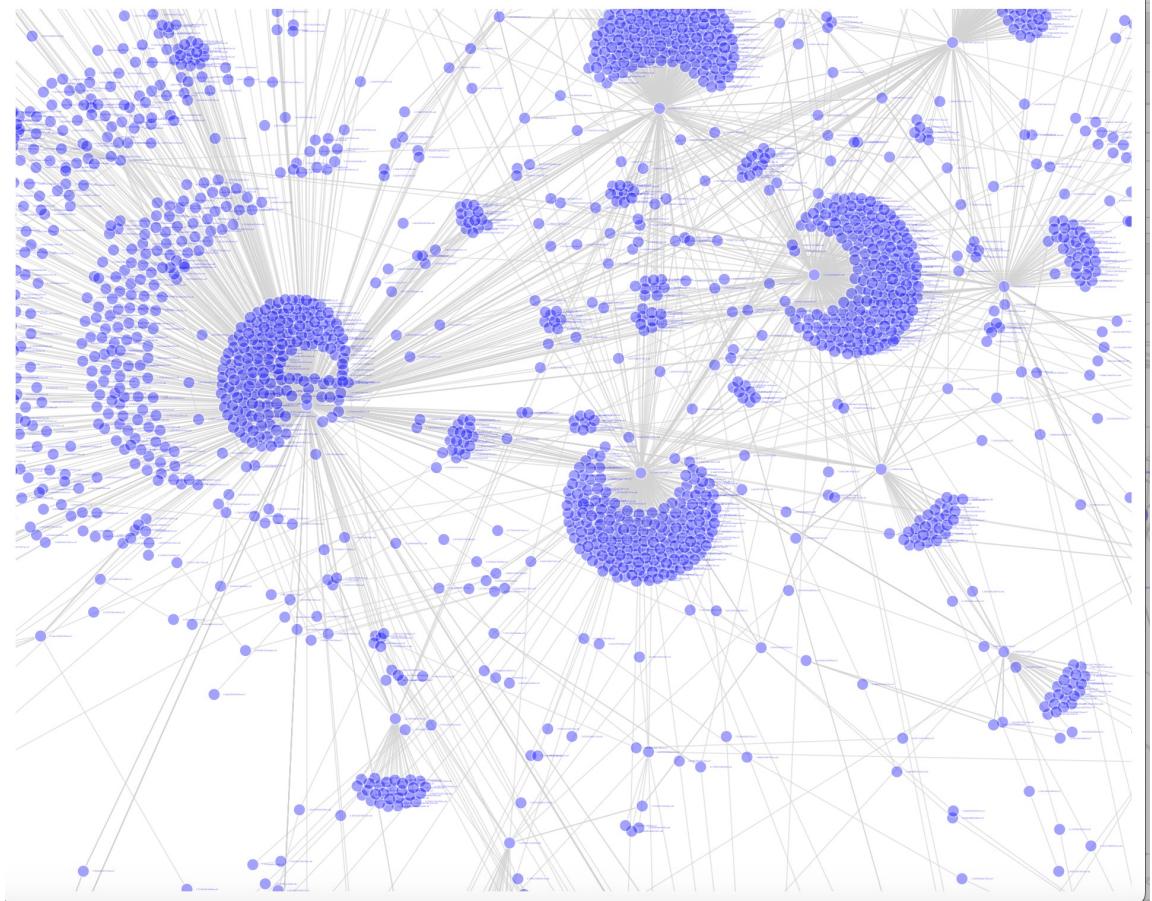


Problem Thinking

- **What is Fraud Detection**
- **Why this Approach**

Sampling

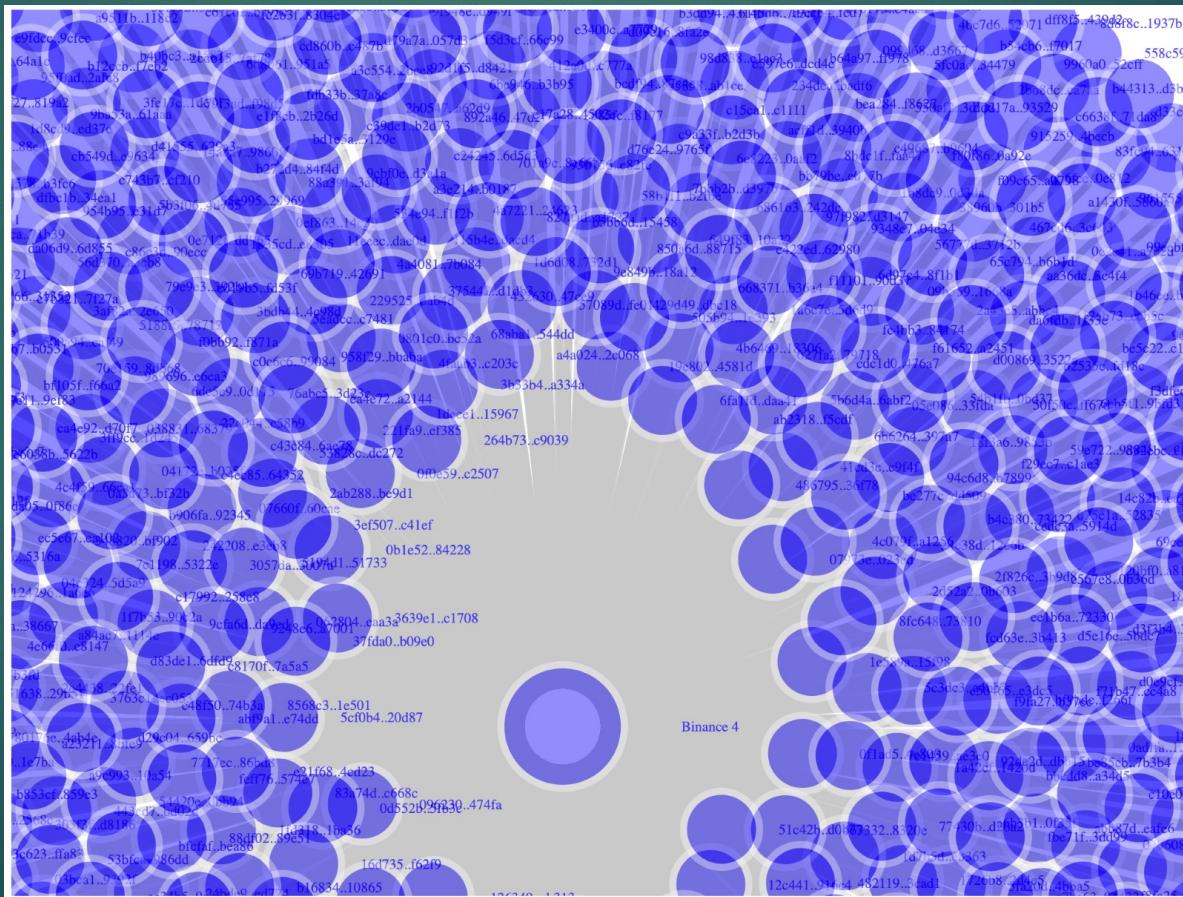
- ▶ Why sample trimming?
- ▶ Approach A
- ▶ largest component
- ▶ Approach B
- ▶ Random Sample



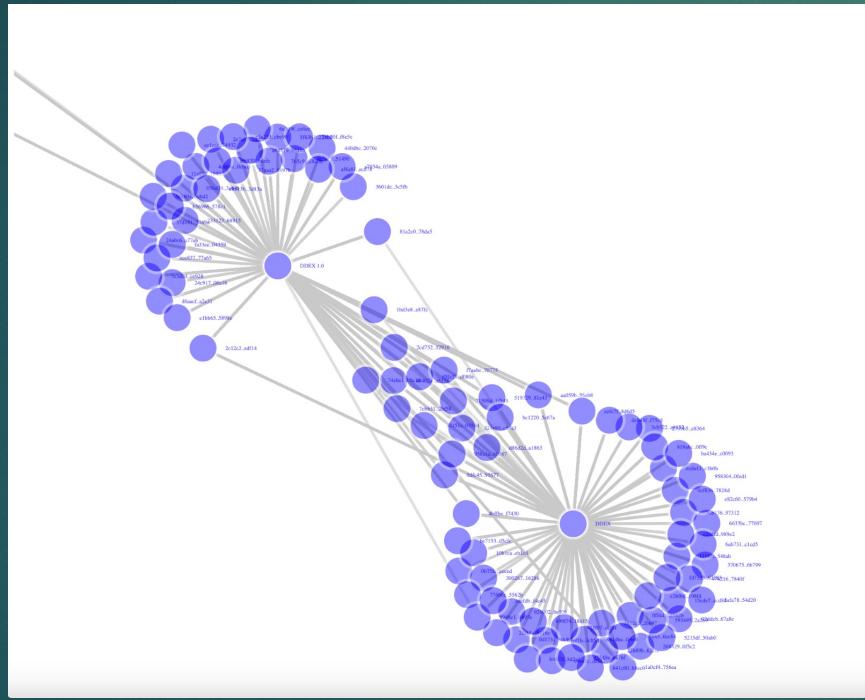
Five Point Summary

- ▶ Size
- ▶ Density
- ▶ Clustering coefficient
- ▶ Degree distribution
- ▶ Diameter

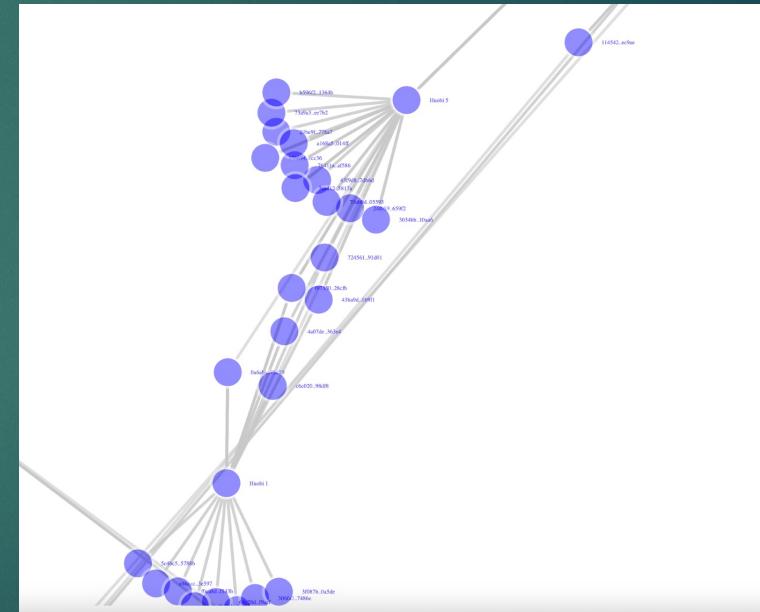
Hotspots: Binance centralized exchange (cex)



Manual Tracking of Money Laundering



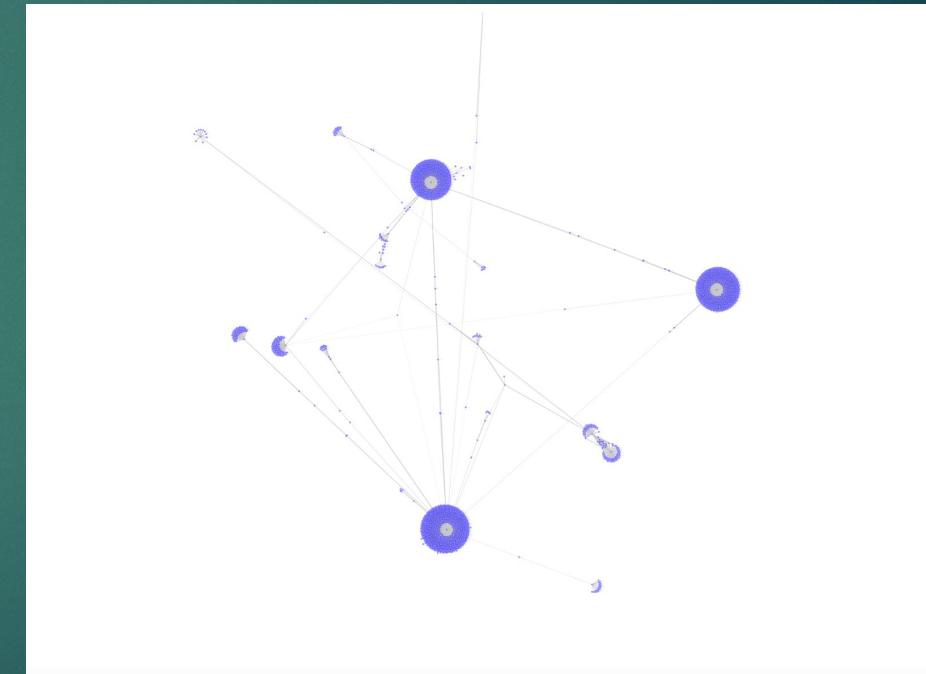
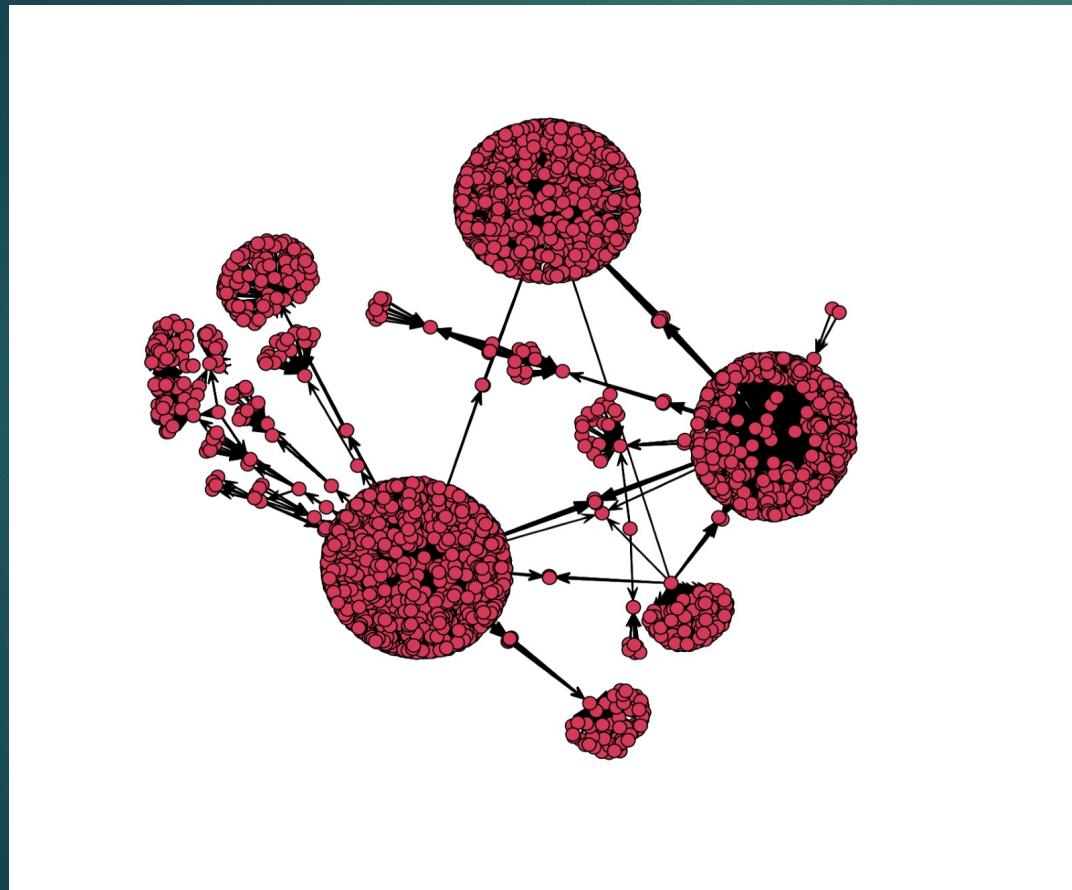
Suspect Transaction



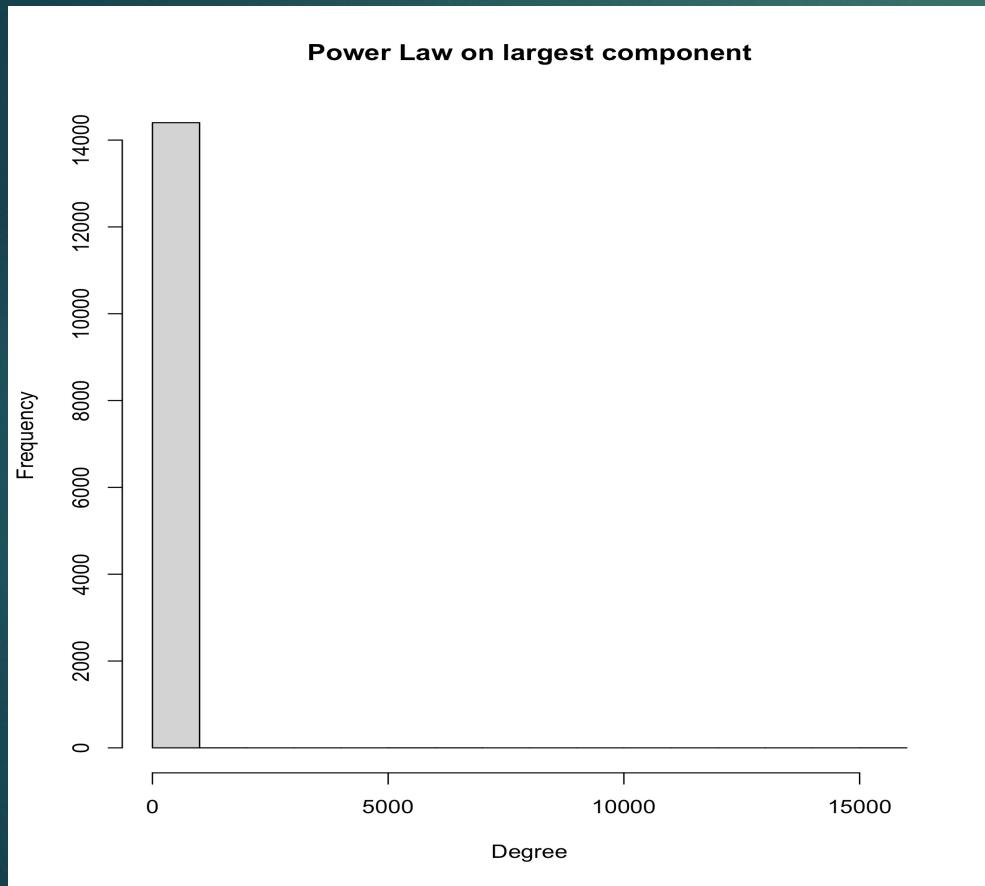
Good Transaction

Approach A

Large Component

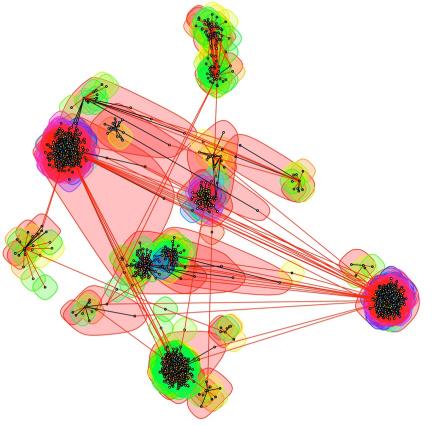


Power Law on largest component

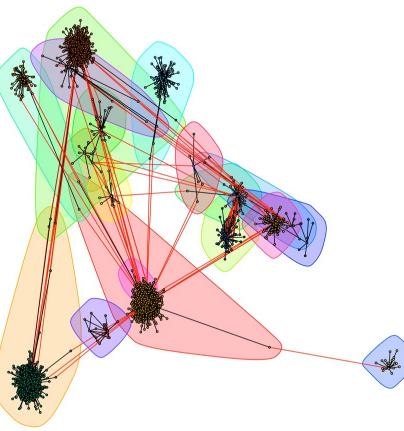


	continuous	alpha	xmin	logLik	KS.stat	KS.p
1	0	2.681	1	-7518.992	0.018	0.007

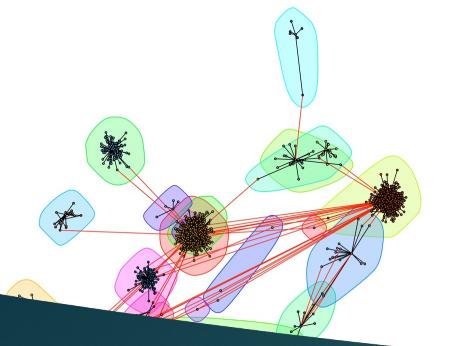
community for cluster_walktrap



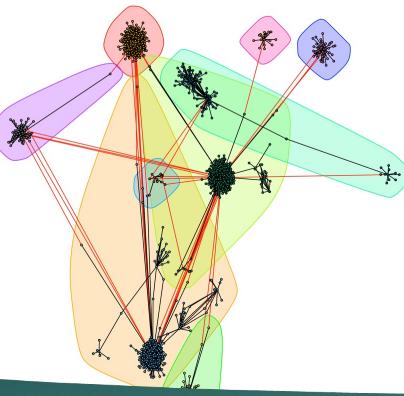
community for cluster_label_prop



community for cluster_infomap



community for cluster_fast_greedy



modularity

<i>cluster_walktrap</i>	0.3082354
<i>cluster_label_prop</i>	0.3084468
<i>cluster_infomap</i>	0.3084454
<i>cluster_fast_greedy</i>	0.3085777
<i>cluster_louvain</i>	0.3085777
<i>cluster_edge_betweenness</i>	0.3085764

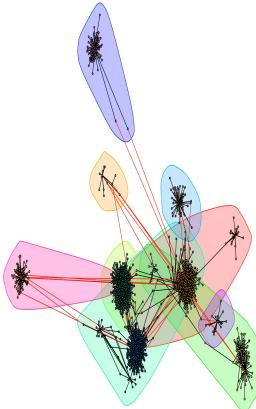
Modularity

Centrality

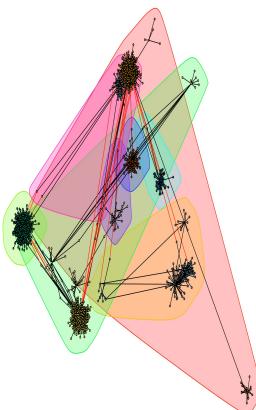
Rand Index comparing various centralities

	cluster_walktrap	cluster_label_prop	cluster_infomap	cluster_fast_greedy	cluster_louvain	cluster_edge_betweenness
cluster_walktrap						
cluster_label_prop	0.09917					
cluster_infomap	0.10003	0.99466				
cluster_fast_greedy	0.09018	0.94736	0.94439			
cluster_louvain	0.09018	0.94736	0.94439	1		
cluster_edge_betweenness	0.08981	0.94338	0.94464	0.98543	0.98543	

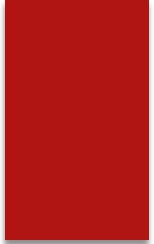
community for cluster_louvain



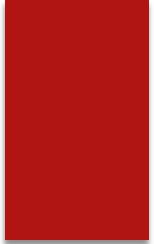
community for cluster_edge_betweenness



Closeness



Jacquard & Vertex Proximity

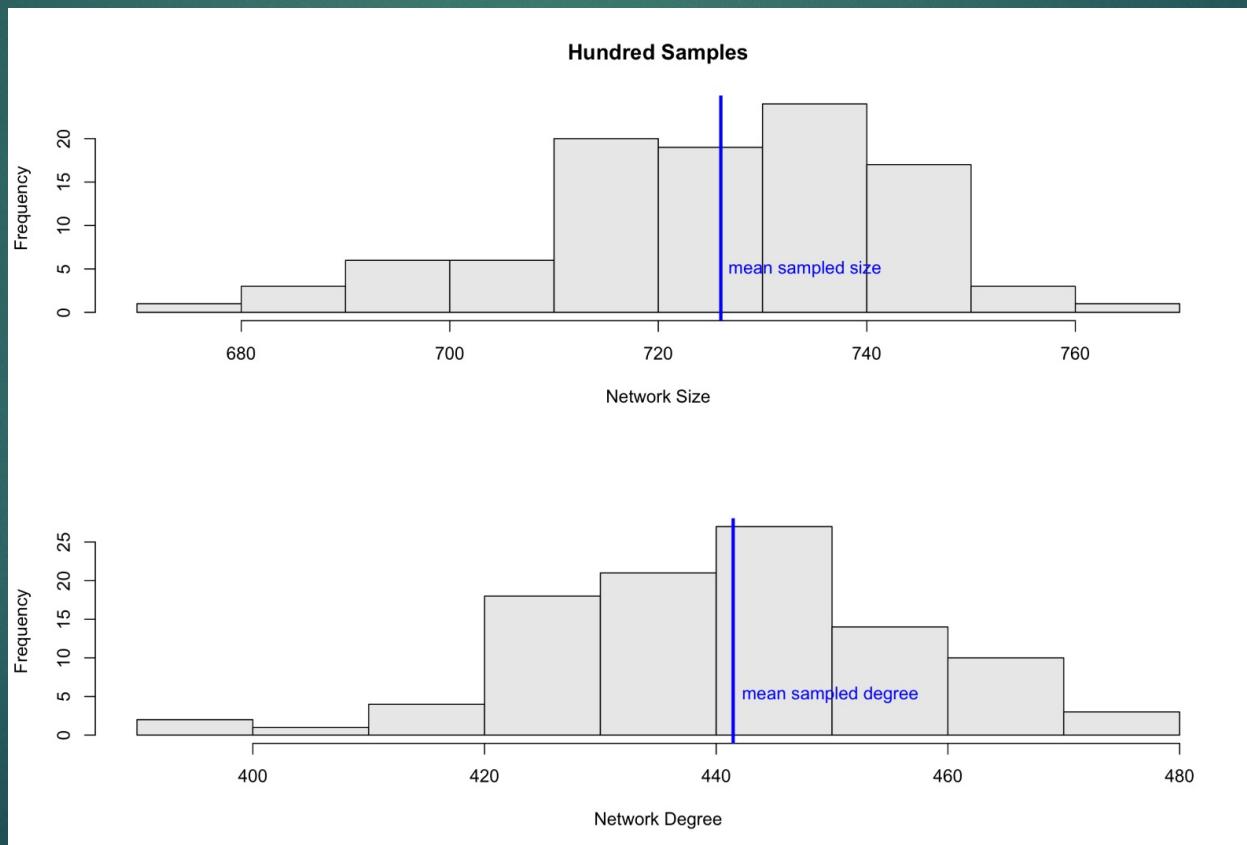


Approach B

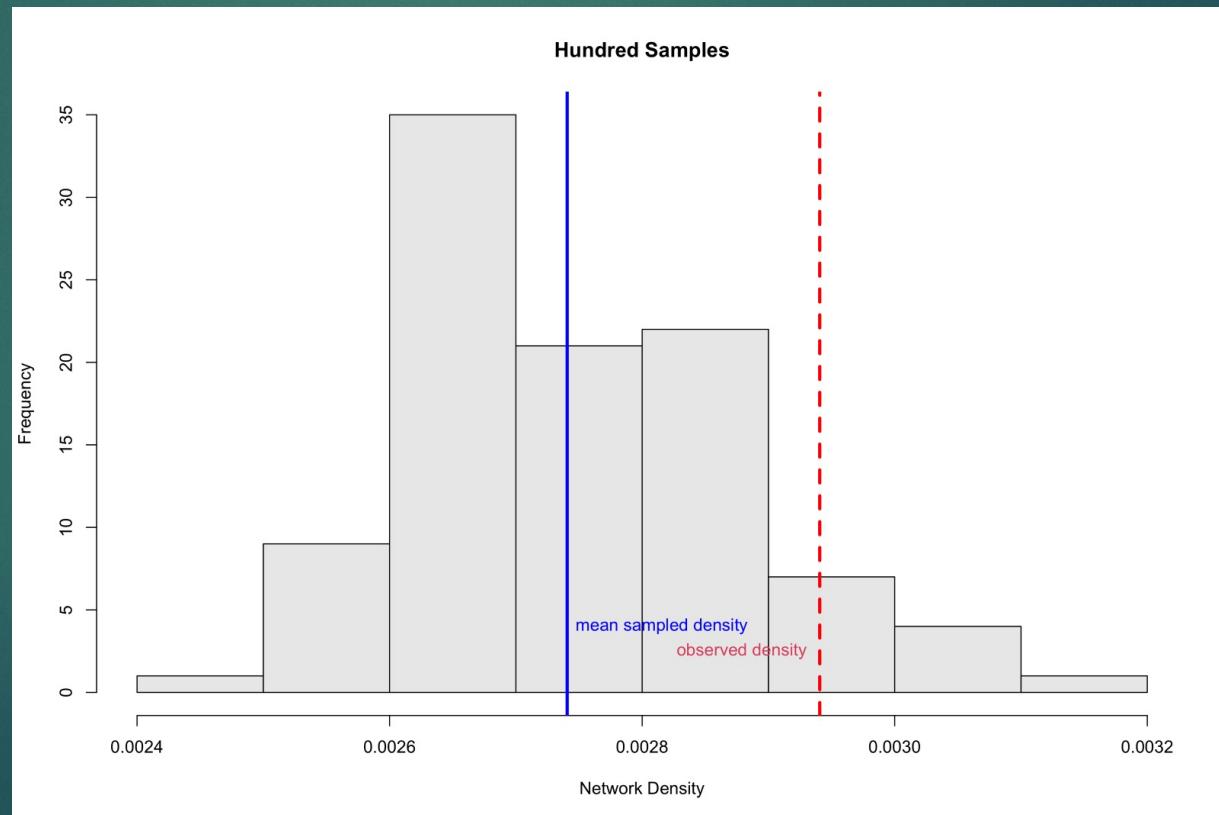
Random Sample

- 10% Random sample

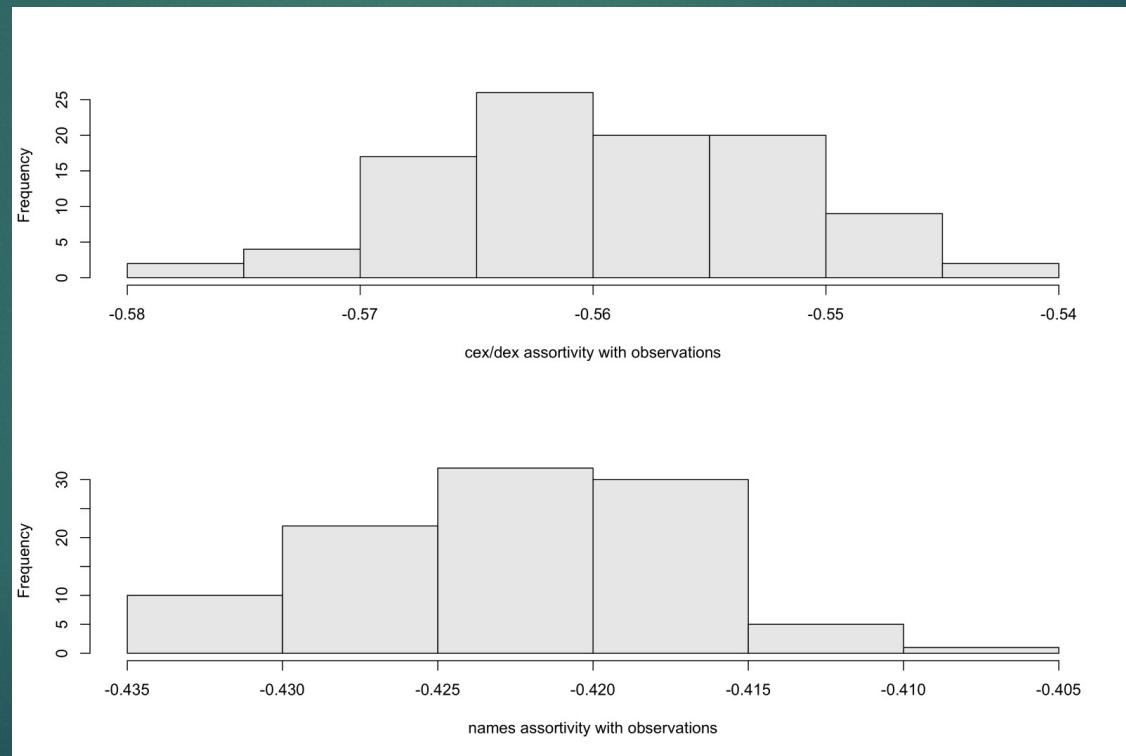
Size & Degree



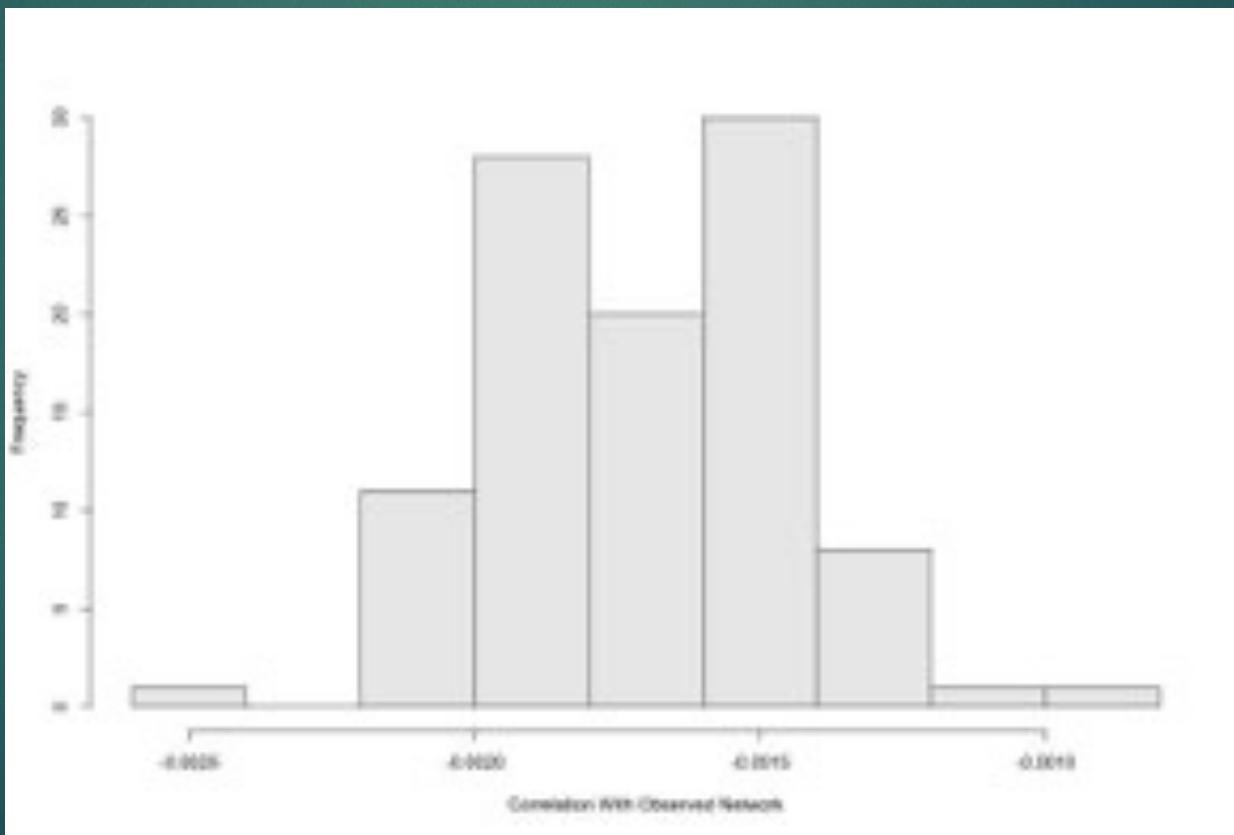
Density



Assortativity



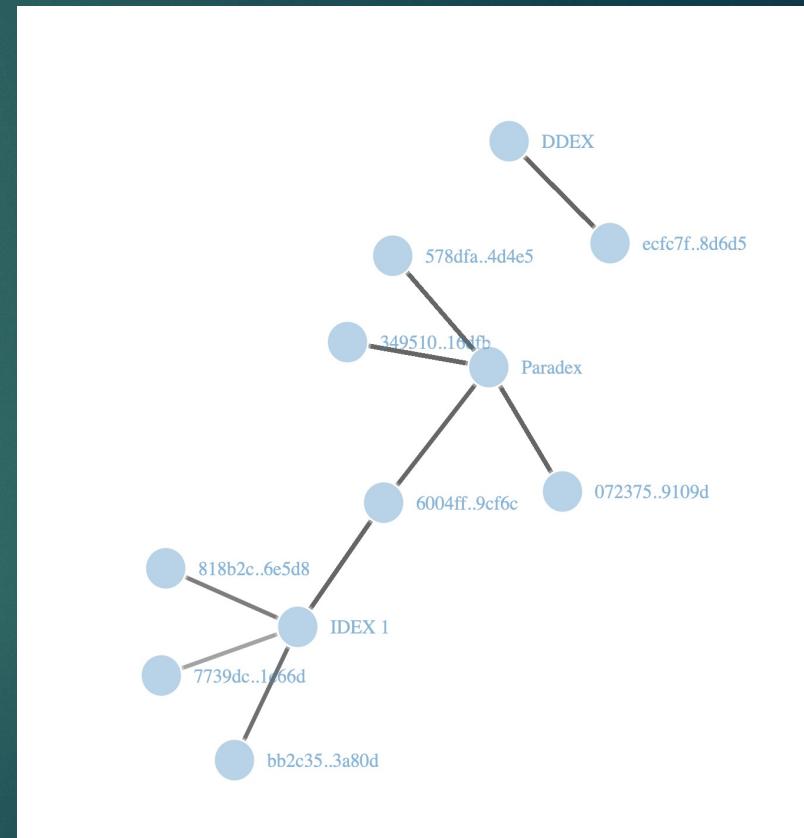
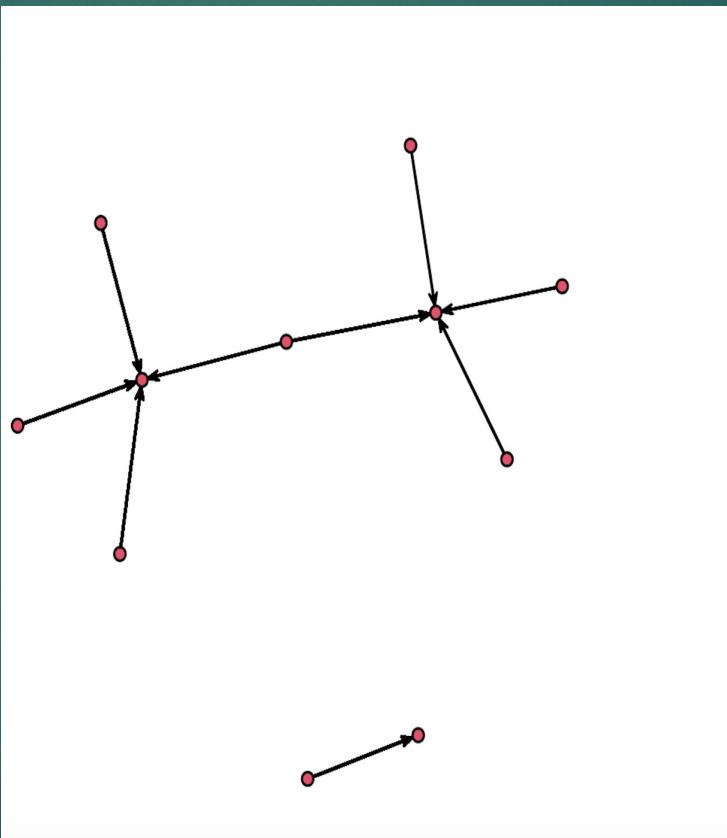
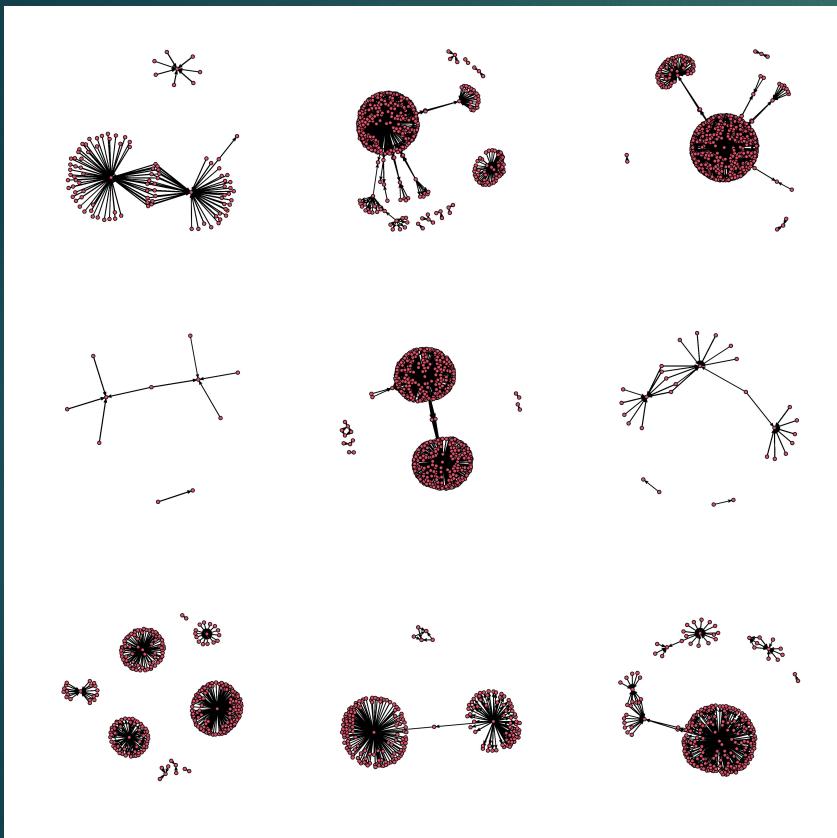
Correlation with other networks



Model : Introduction

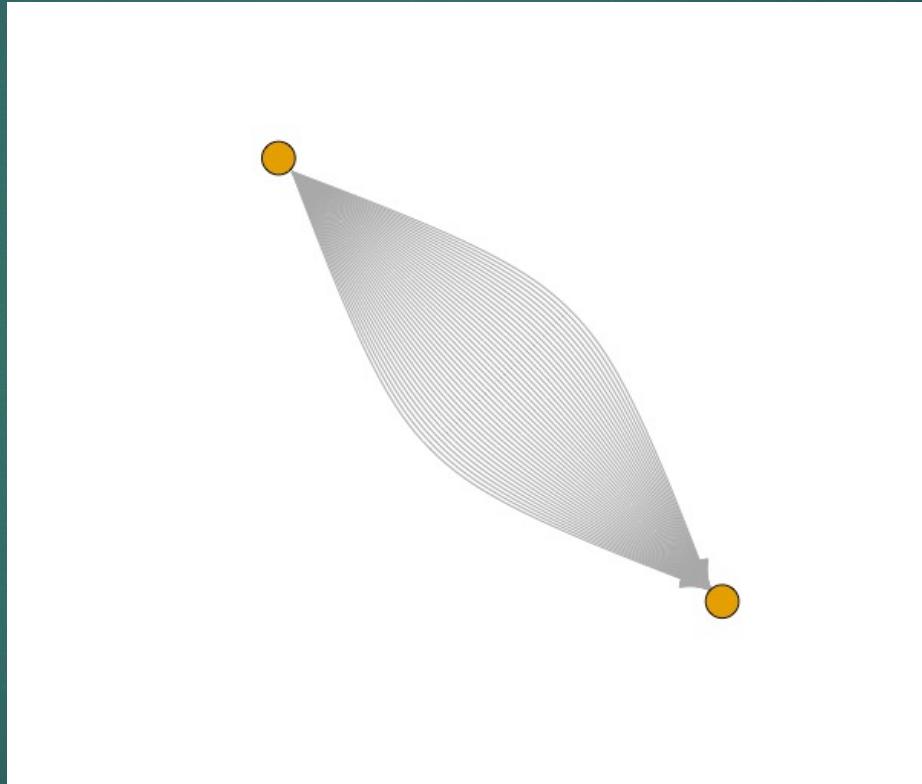
Top 8 Exchanges		Top 20 Tokens		
Exchange	Num Txns	Tokens	Num Txns	
1	2776	1 0xc02aaa39b223fe8d0a0e5c4f27ead9083c756cc2	19991	
2 Huobi	4	2 0x0d8775f648430679a709e98d2b0cb6250d2887ef	3679	
3 Binance	2	3 0x0000000000085d4780b73119b644ae5ecd22b376	3382	
4 Bitmax	2	4 0x9f8f72aa9304c8b593d555f12ef6589cc3a579a2	1680	
5 DDEX	2	5 0xb8c77482e45f1f44de1745f52c74426c631bdd52	1614	
6 Gate	2	6 0x89d24a6b4ccb1b6faa2625fe562bdd9a23260359	1368	
7 ABCC	1	7 0xa15c7ebe1f07caf6bff097d8a589fb8ac49ae5b3	989	
8 Bittrex	1	8 0xa0b86991c6218b36c1d19d4a2e9eb0ce3606eb48	727	
Exchange Types		9 0x6f259637dcd74c767781e37bc6133cd6a68aa161	635	
		10 0xe41d2489571d322189246dafa5ebde1f4699f498	446	
		11 0xd26114cd6ee289accf82350c8d8487fdb8a0c07	363	
		12 0x8971f9fd7196e5cee2c1032b50f656855af7dd26	356	
		13 0xf629cbd94d3791c9250152bd8dfbdf380e2a3b9c	251	
		14 0x8e870d67f660d95d5be530380d0ec0bd388289e1	237	
		15 0xdd974d5c2e2928dea5f71b9825b8b646686bd200	220	
		16 0x0f5d2fb29fb7d3cfbee44a200298f468908cc942	200	
		17 0xb64ef51c888972c908cfacf59b47c1afbc0ab8ac	80	
		18 0x1f573d6fb3f13d689ff844b4ce37794d79a7ff1c	78	
		19 0x514910771af9ca656af840dff83e8264ecf986ca	66	
		20 0x4dc3643dbc642b72c158e7f3d2ff232df61cb6ce	41	
Component summary				
size	density	diameter	cl.coeff	txns
1 2799	0.0046482	9.5147495923e+25	0.0080961	36403

Top 9 Transaction:



Heavy Flow model

23



Base Model:

```

Call:
ergm(formula = train.nw.inw ~ edges)

Maximum Likelihood Results:

  Estimate Std. Error MCMC % z value Pr(>|z|)
edges -7.21808   0.01868     0 -386.5 <1e-04 ***
---
Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1

  Null Deviance: 5428453  on 3915801  degrees of freedom
  Residual Deviance: 47157  on 3915800  degrees of freedom

AIC: 47159  BIC: 47173  (Smaller is better. MC Std. Err. = 0)
-----
probability
  edges
0.0007326726

Call:
ergm(formula = train.nw.inw ~ edges + nodefactor("type"))

Maximum Likelihood Results:

  Estimate Std. Error MCMC % z value Pr(>|z|)
edges      -9.80102  0.06592     0 -148.67 <1e-04 ***
nodefactor.type.cex 6.86236  0.06868     0  99.92 <1e-04 ***
nodefactor.type.dex 4.26954  0.11552     0  36.96 <1e-04 ***
---
Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1

  Null Deviance: 5428453  on 3915801  degrees of freedom
  Residual Deviance: 24427  on 3915798  degrees of freedom

AIC: 24433  BIC: 24472  (Smaller is better. MC Std. Err. = 0)
-----
probability
  edges nodefactor.type.cex nodefactor.type.dex
5.539217e-05      9.989547e-01      9.862048e-01

```

```

Maximum Likelihood Results:

  Estimate Std. Error MCMC % z value Pr(>|z|)
edges      -9.80102  0.06592     0 -148.67 <1e-04 ***
nodefactor.type.cex 6.86236  0.06868     0  99.92 <1e-04 ***
nodefactor.type.dex 4.26954  0.11552     0  36.96 <1e-04 ***
nodematch.address    -Inf  0.00000     0  -Inf <1e-04 ***
---
Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1

  Null Deviance: 5428453  on 3915801  degrees of freedom
  Residual Deviance: 24427  on 3915797  degrees of freedom

AIC: 24433  BIC: 24472  (Smaller is better. MC Std. Err. = 0)

Warning: The following terms have infinite coefficient estimates:
  nodematch.address
-----
probability
  edges nodefactor.type.cex nodefactor.type.dex  nodematch.address
5.539217e-05      9.989547e-01      9.862048e-01      0.000000e+00

Call:
ergm(formula = train.nw.inw ~ edges + nodefactor("type") + nodematch("name2"))

Maximum Likelihood Results:

  Estimate Std. Error MCMC % z value Pr(>|z|)
edges      -2.2761  0.4561     0 -4.991 <1e-04 ***
nodefactor.type.cex -0.5218  0.4549     0 -1.147  0.251
nodefactor.type.dex -2.4086  0.4615     0 -5.219 <1e-04 ***
nodematch.name2     -12.8879 1.0990     0 -11.727 <1e-04 ***
---
Signif. codes:  0 '****' 0.001 '***' 0.01 '**' 0.05 '.' 0.1 ' ' 1

  Null Deviance: 5428453  on 3915801  degrees of freedom
  Residual Deviance: 22563  on 3915797  degrees of freedom

AIC: 22571  BIC: 22624  (Smaller is better. MC Std. Err. = 0)
-----
probability
  edges nodefactor.type.cex nodefactor.type.dex  nodematch.name2
9.312236e-02      3.724277e-01      8.251967e-02      2.528372e-06

```



QUESTIONS?

THANK YOU