

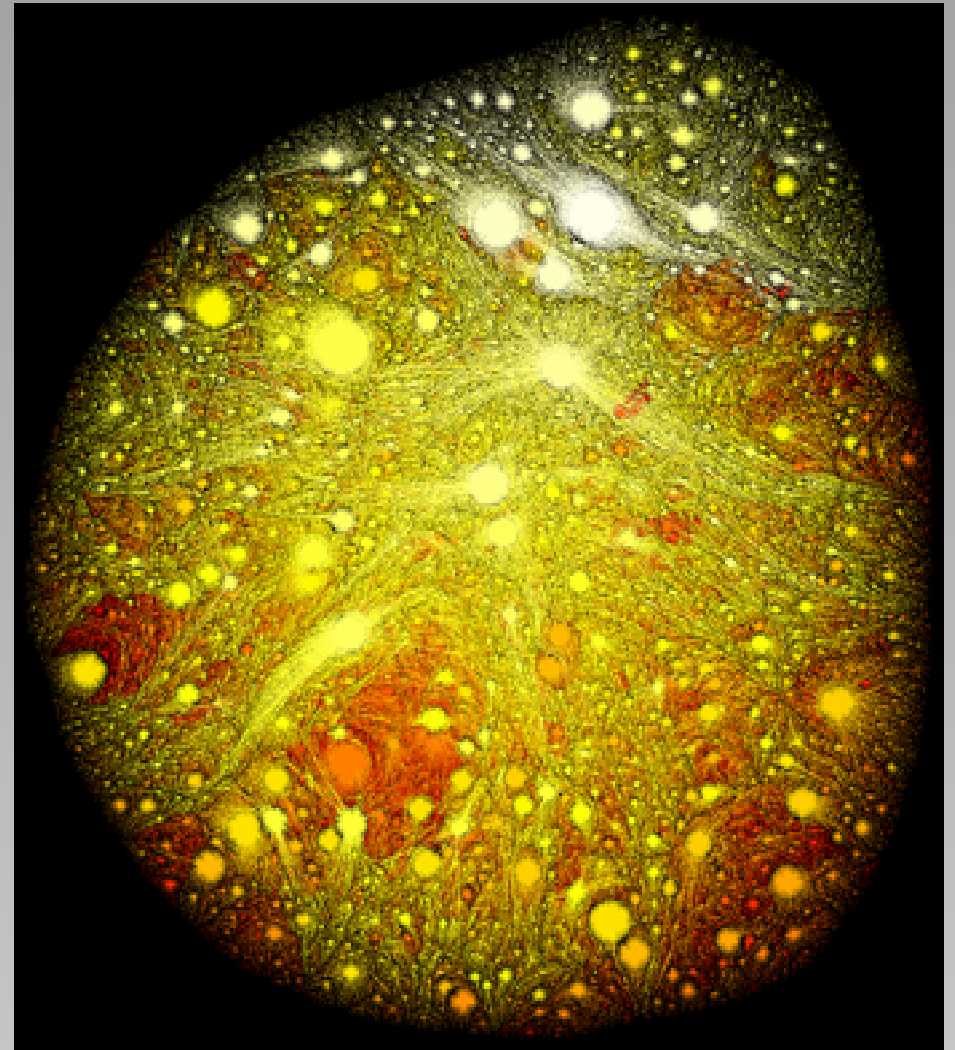
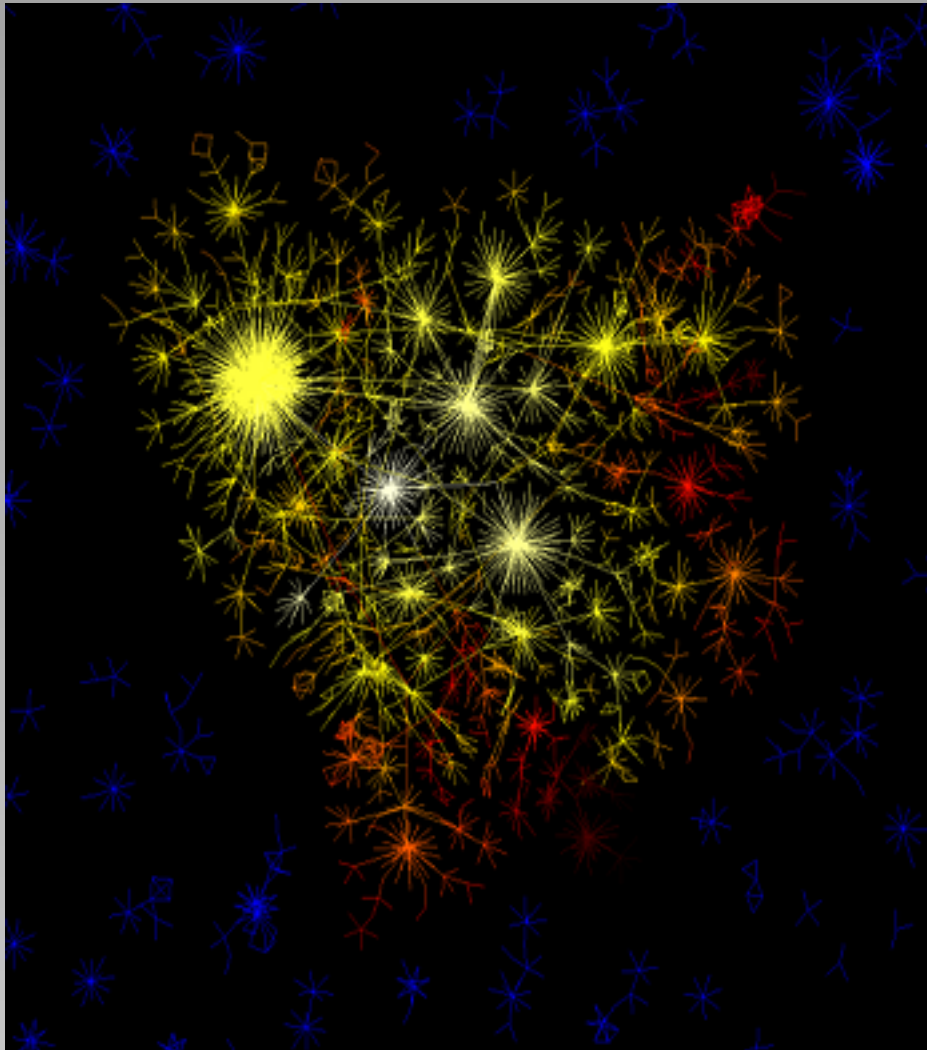
Lossy Matrix Compression

Or: How I Stopped Worrying About Network Visualization and Started Visualizing Networks

Erich Morisse
@emorisse
HowWeKnowUs.com

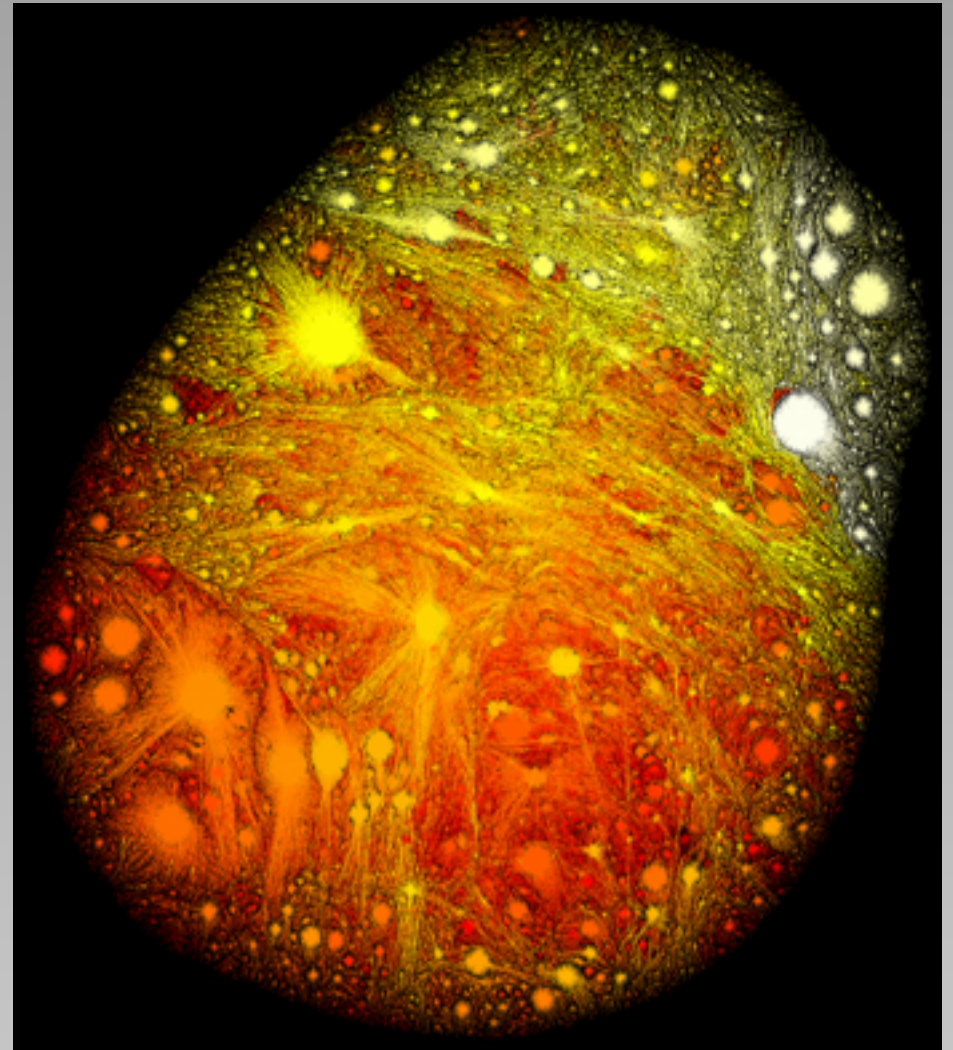


Network Clarity Drops with Scale

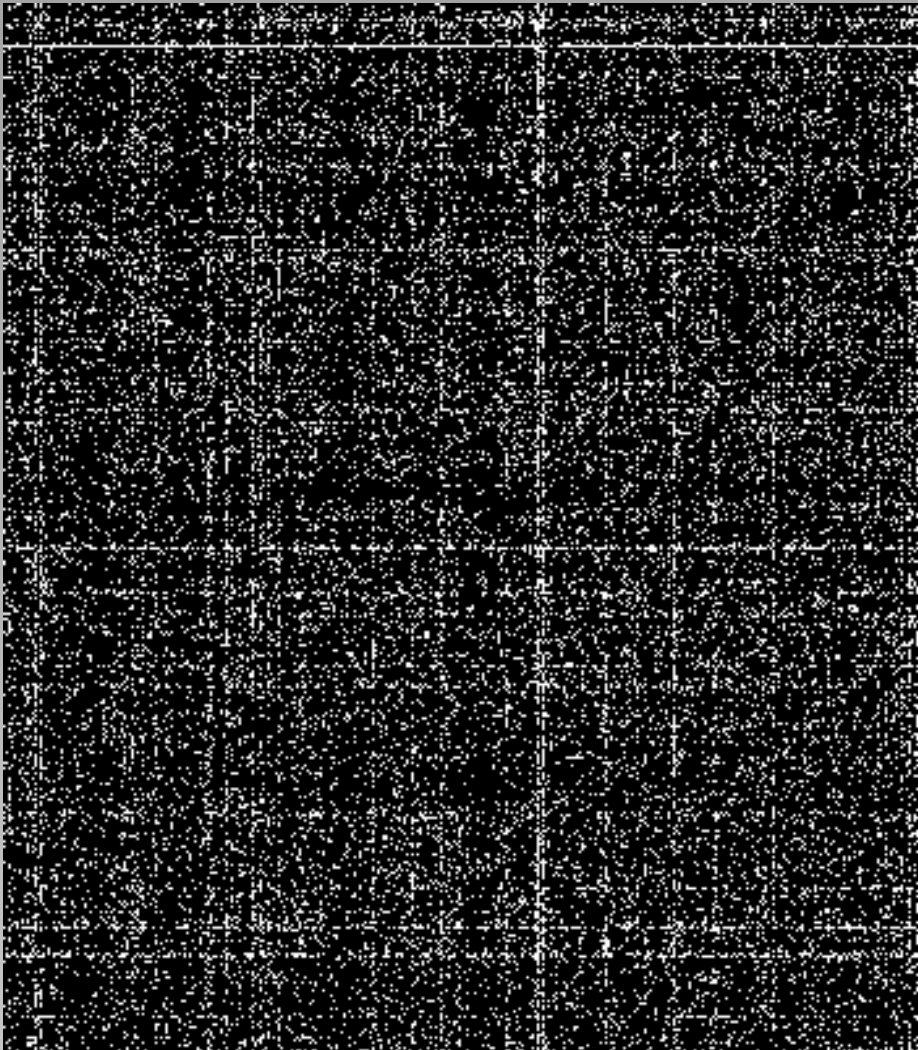


Processing Requirements are N^2

- 1 week for 1,200,000 people
- But, I want 64,000,000 people
- $64/1.2 = 50$
- $50^2 = 2500$ Weeks
- I'm not a patient man



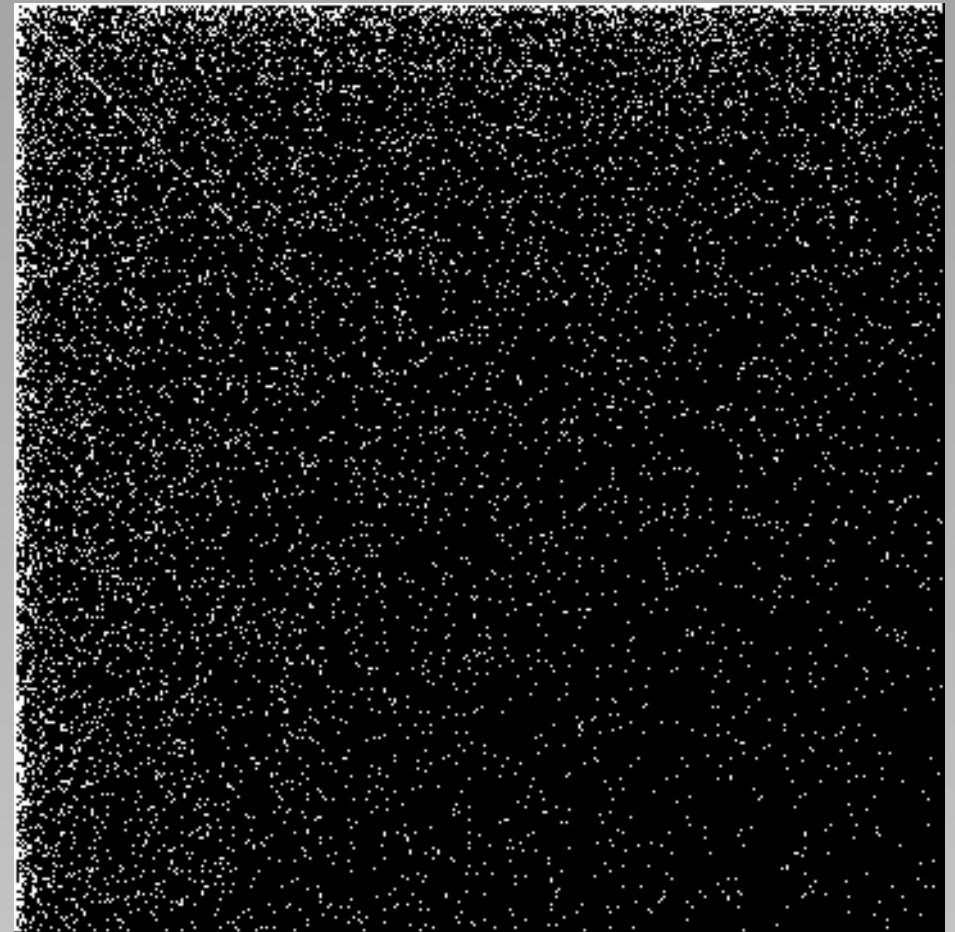
Enter “Lossy Compressed” Matrix



- User Y knows User X
- Collision on Y (or X) changes color pixel
- Recoverable (identifiable users) compression to 32:1
- Non-recoverable to over $2^{32}:1$
- Non- or Directed

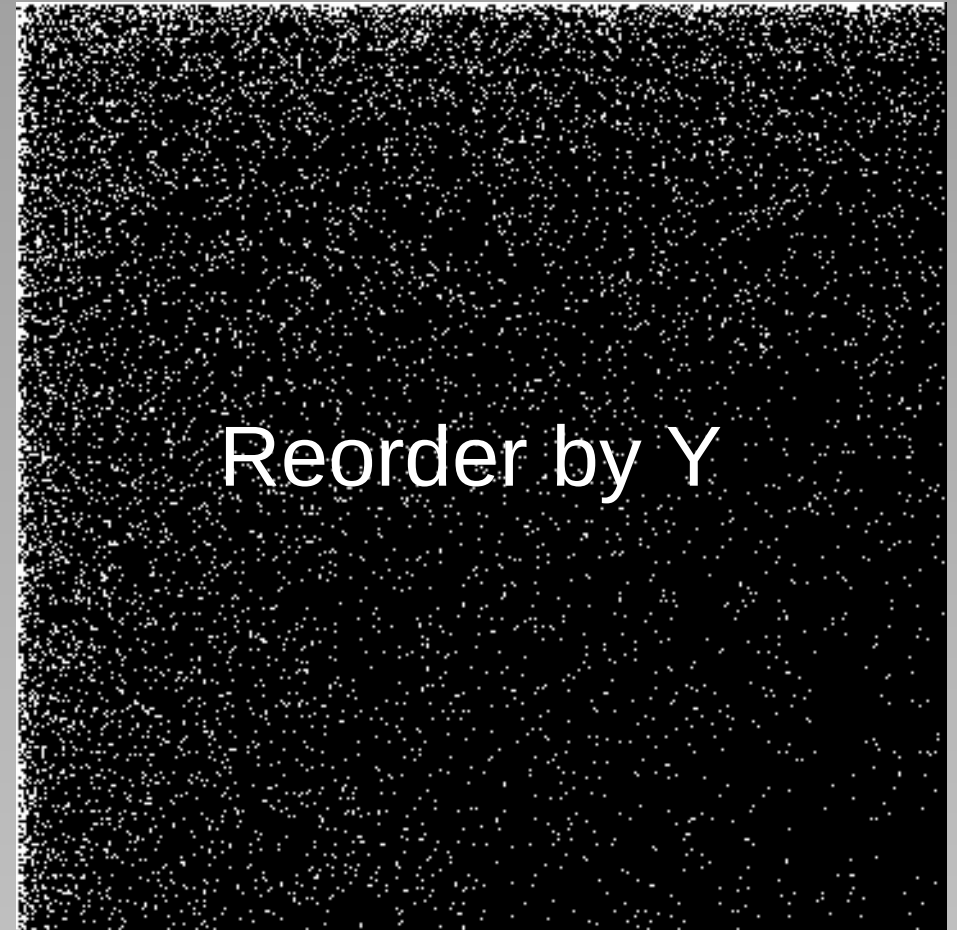
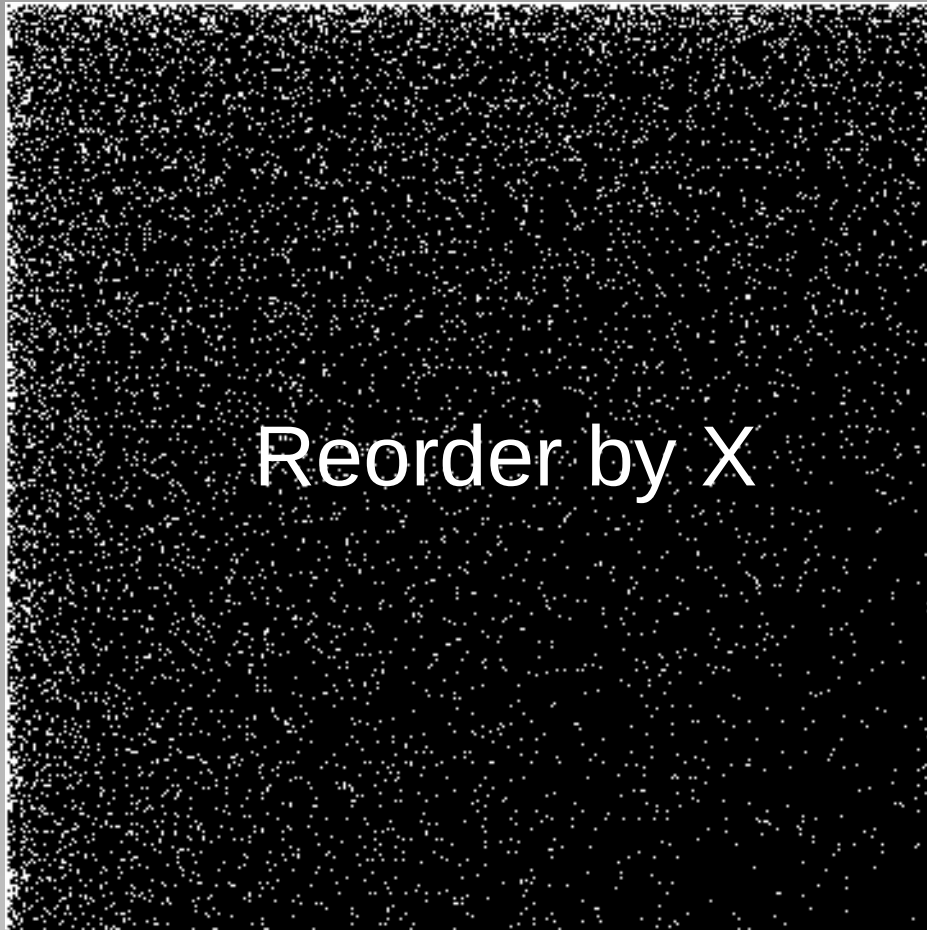
Allows Streaming Network Analysis

- First seen node = 0, next node = 1, etc.
- Can resize matrix on the fly
- Downside: creates artificial “near-identity” line



1.2M Brand conversations on Twitter

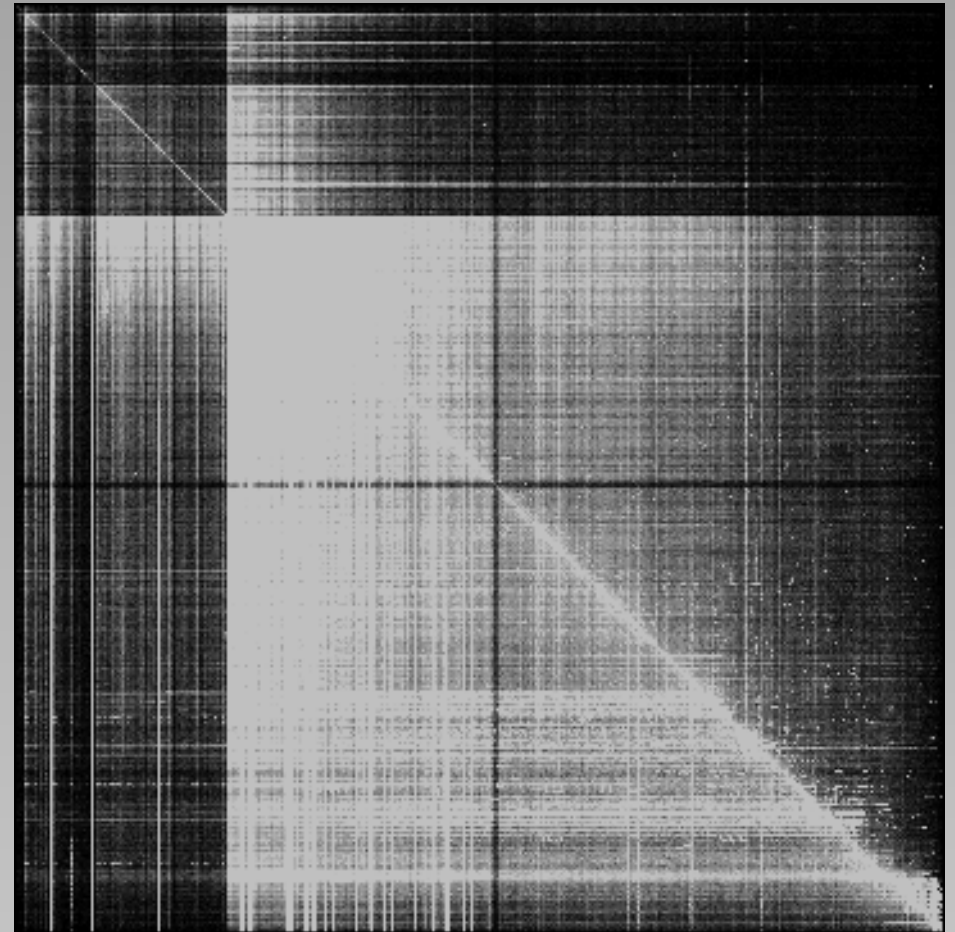
New Option: Test for “Balance”



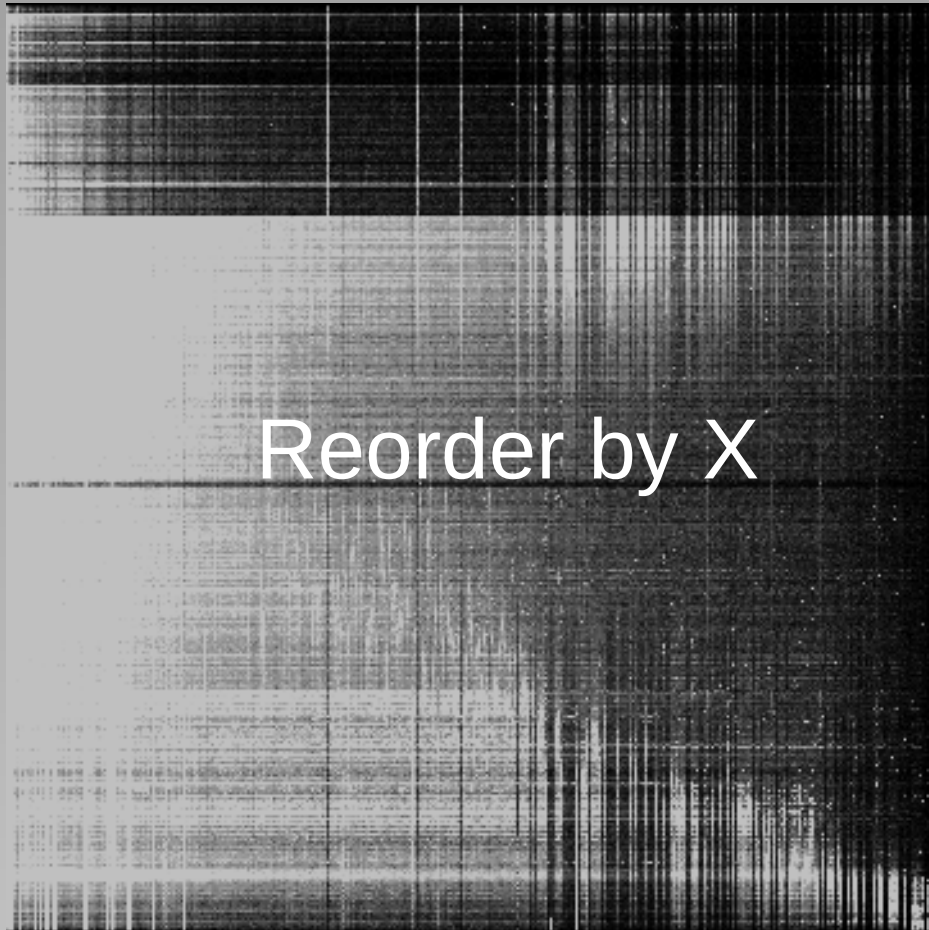
Nice Side Effect: Drops false near-identity line

New Option: Add Temporality

- Add exogenous person identifiers incremented on creation date
- Tell the story of the evolving network
- 64,000,000 People
- 1,450,000,000 Relationships



Example: Test for “Balance”



Clearly a non-isometric, directed graph

What's the Story?

Product begins

Beta

Open to Friends

Private Beta

Open to Public

Rapid Growth

User ID change?

Celebrity PR
machines join?

Content harvesting
goes big-time

Early Users have
high reciprocity in
follows

Users follow more
users joining
before they did

New users bring
new users (fuzzy
identity line)

Large drag of bots
on the System

Bot controls
installed

What Else Can We See?

Impact of Lossy Matrix Visualization

- 2625.5:1 compression over *sparse* graph
- Visualization suitable for analysis in hours, not weeks or years
- Legibility at large scale
- Processing requires laptop, not supercomputer

