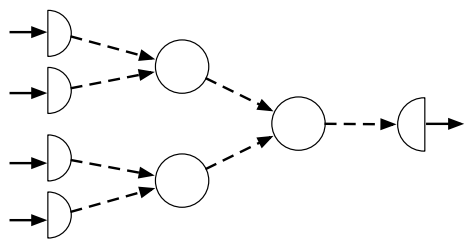
Raphtory: A new tool for large temporal networks applied to the far right social network Gab



Presenter: Imane Hafnaoui,
Project team (alphabetical order):
Naomi Arnold, Richard G. Clegg, Félix Cuadrado, Raul Mondragon, Hugo Parada, Ben Steer

Aims of this talk (unordered list)

Aim A: Give insight into data

What drives the evolution of the alt-right social network gab?

Aims of this talk (unordered list)

Aim A: Give insight into data

What drives the evolution of the alt-right social network gab?

Aim 1: Show a technique

Temporal analysis and in particular "windowing" is an invaluable tool for social network analysis.

Aims of this talk (unordered list)

Aim A: Give insight into data

What drives the evolution of the alt-right social network gab?

Aim 1: Show a technique

Temporal analysis and in particular "windowing" is an invaluable tool for social network analysis.

Aim α: Sell you a tool

Raphtory is an open-source, big-data platform developed at QMUL. It is unique in its ability to perform flexible temporal analysis on batch or streamed graph data.

The gab platform (and why we care)

- 95GB raw data (19 million posts) from gab platform (medium data).
  - Data is user, posts (threaded), timestamps and other metadata.
  - NB our research on structure not content.
- First eighteen months of data available.
  - September 2016 – May 2018.
- Largely complete data from this period.
  - "free speech" focus means "everything" is public.
- Alt-right focus:
  - Racism, fake news, hate speech, radicalisation.

The gab platform (and why we care)

- 95GB raw data (19 million p… … data).
  - Data is user, posts (threaded) … mps …
  - NB our research on s… …nt.
- First eighteen m…
  - S…
- … this period.
- … focus means "everything" is public.
- … focus:
  - Racism, fake news, hate speech, radicalisation.

The gab platform (and why we care)

- 95GB data (19 million p… data).
  - posts (threaded) … mps a…
  - NB … n s… nt.
- First eighteen …
  - S…
- …d.
- … focus means … public.
- … focus:
  - Racism, fake news, hate speech, radicali…

The gab platform (and why we care)

- 95GB data (19 million posts (threaded)
  - NB
- First eighteen
  - Sc
- public.
- Focus:
  - Racism, fake ne peech, radicali

Temporal networks

Social network $\rightarrow$ graph G

We can form a graph from a social network by creating an edge between two users if they have a relationship, interact, etc.

Temporal networks

Social network $\rightarrow$ graph G

We can form a graph from a social network by creating an edge between two users if they have a relationship, interact, etc.

Social network $\rightarrow$ many graphs $G(t, \tau)$

There is no such thing as "the twitter graph". The graph varies hugely depending on the time t and timescale τ.

Temporal networks

Social network $\longrightarrow$ graph G

We can form a graph from a social network by creating an edge between two users if they have a relationship, interact, etc.

Social network $\longrightarrow$ many graphs $G(t, \tau)$

There is no such thing as "the twitter graph". The graph varies hugely depending on the time t and timescale τ.





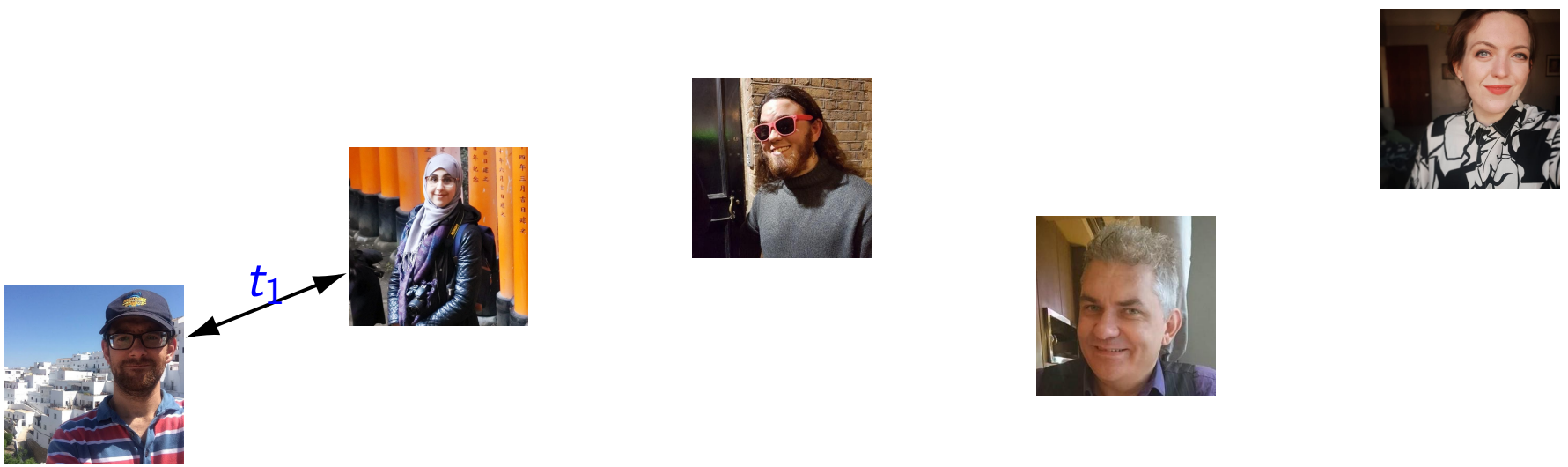
Temporal networks

Social network $\longrightarrow$ graph G

We can form a graph from a social network by creating an edge between two users if they have a relationship, interact, etc.

Social network $\longrightarrow$ many graphs $G(t, \tau)$

There is no such thing as "the twitter graph". The graph varies hugely depending on the time t and timescale τ.



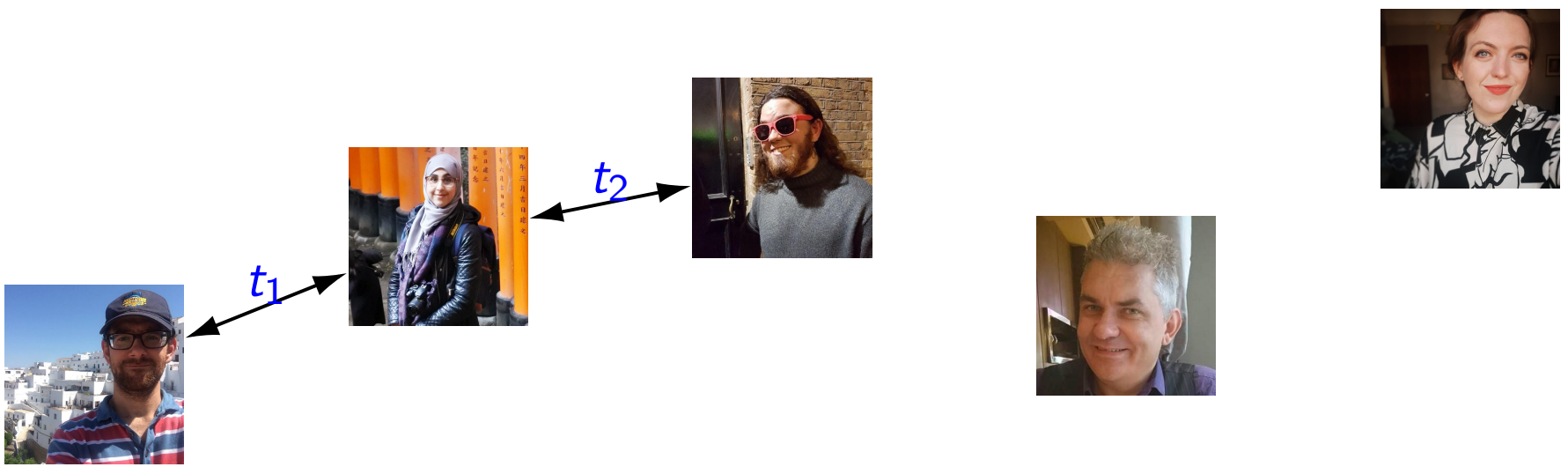
Temporal networks

Social network $\longrightarrow$ graph G

We can form a graph from a social network by creating an edge between two users if they have a relationship, interact, etc.

Social network $\longrightarrow$ many graphs $G(t, \tau)$

There is no such thing as "the twitter graph". The graph varies hugely depending on the time t and timescale τ.



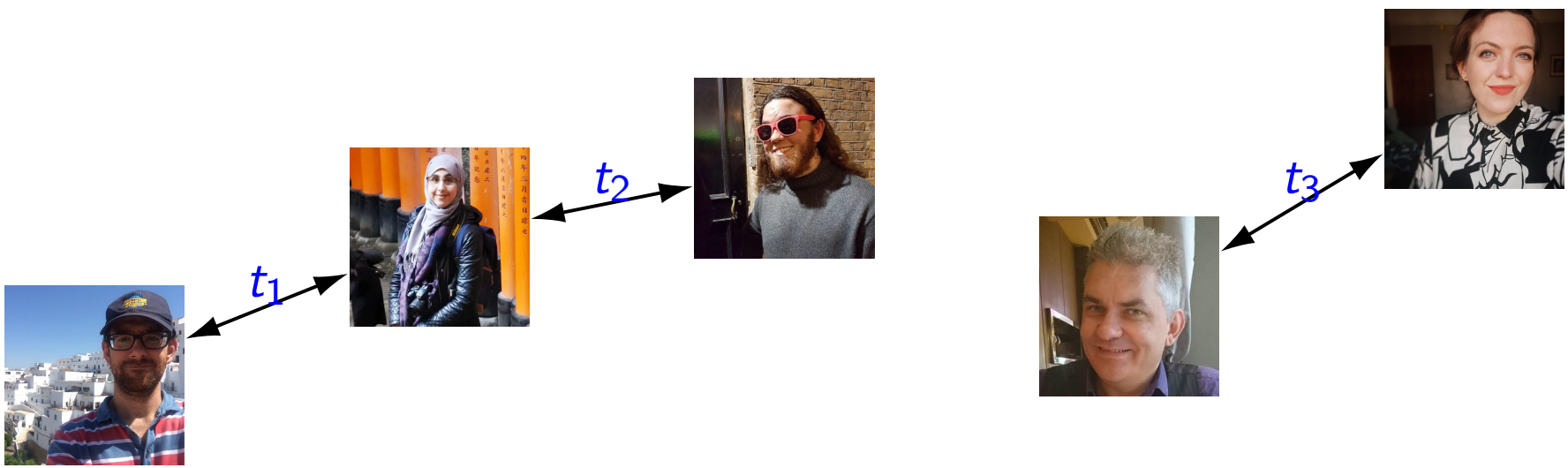
Temporal networks

Social network $\longrightarrow$ graph G

We can form a graph from a social network by creating an edge between two users if they have a relationship, interact, etc.

Social network $\longrightarrow$ many graphs $G(t, \tau)$

There is no such thing as "the twitter graph". The graph varies hugely depending on the time t and timescale τ.



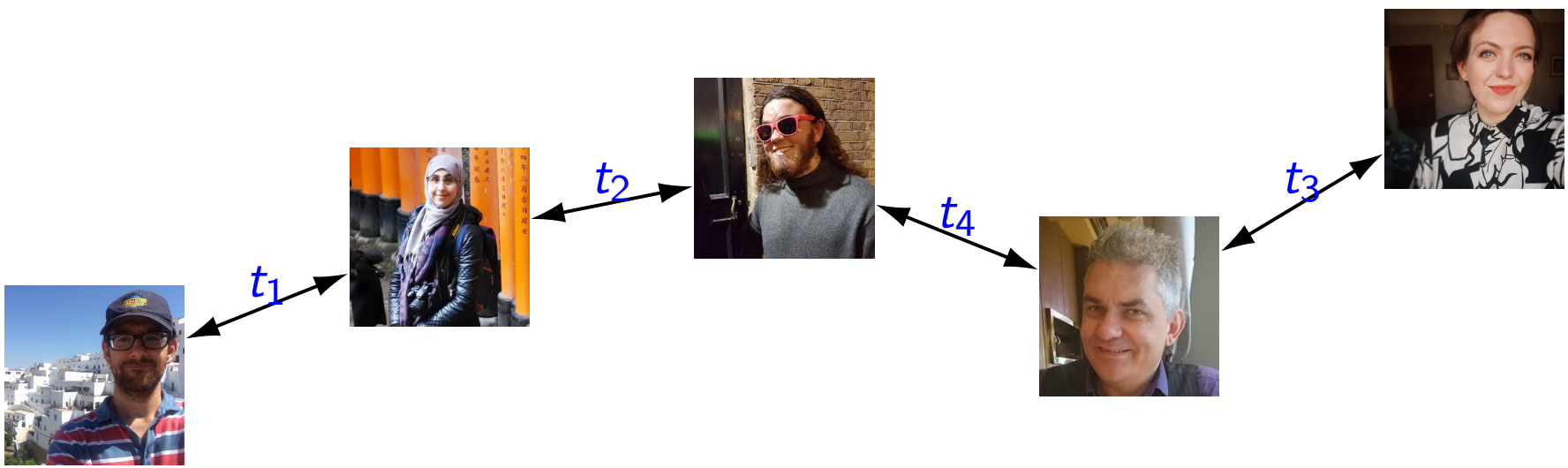
Temporal networks

Social network $\longrightarrow$ graph G

We can form a graph from a social network by creating an edge between two users if they have a relationship, interact, etc.

Social network $\longrightarrow$ many graphs $G(t, \tau)$

There is no such thing as "the twitter graph". The graph varies hugely depending on the time t and timescale τ.



Temporal networks

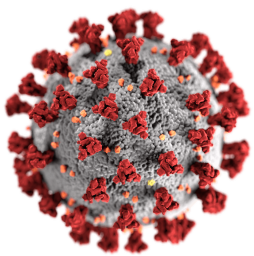
Social network $\longrightarrow$ graph G

We can form a graph from a social network by creating an edge between two users if they have a relationship, interact, etc.

Social network $\longrightarrow$ many graphs $G(t, \tau)$

There is no such thing as "the twitter graph". The graph varies hugely depending on the time t and timescale τ.



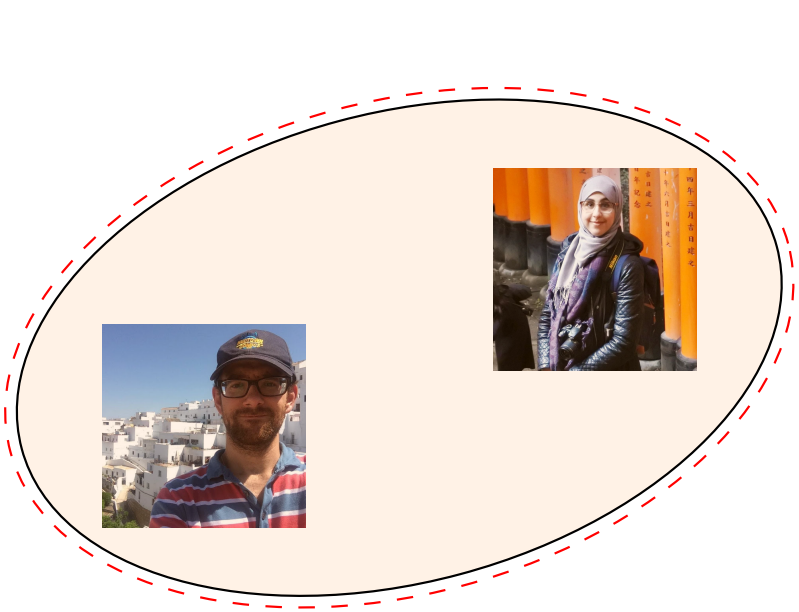

Temporal networks

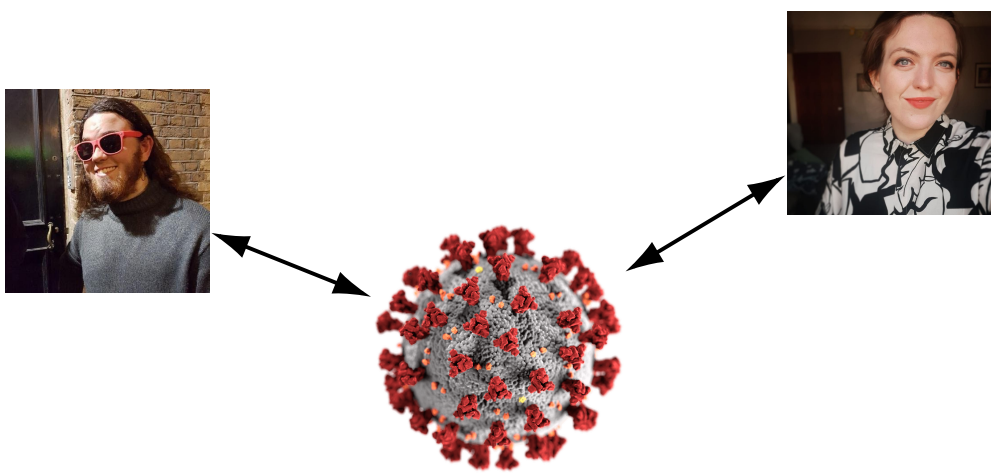
Social network $\rightarrow$ graph G

We can form a graph from a social network by creating an edge between two users if they have a relationship, interact, etc.

Social network $\rightarrow$ many graphs $G(t, \tau)$

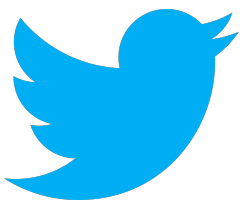
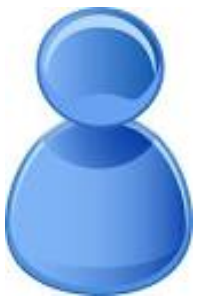
There is no such thing as "the twitter graph". The graph varies hugely depending on the time t and timescale τ.



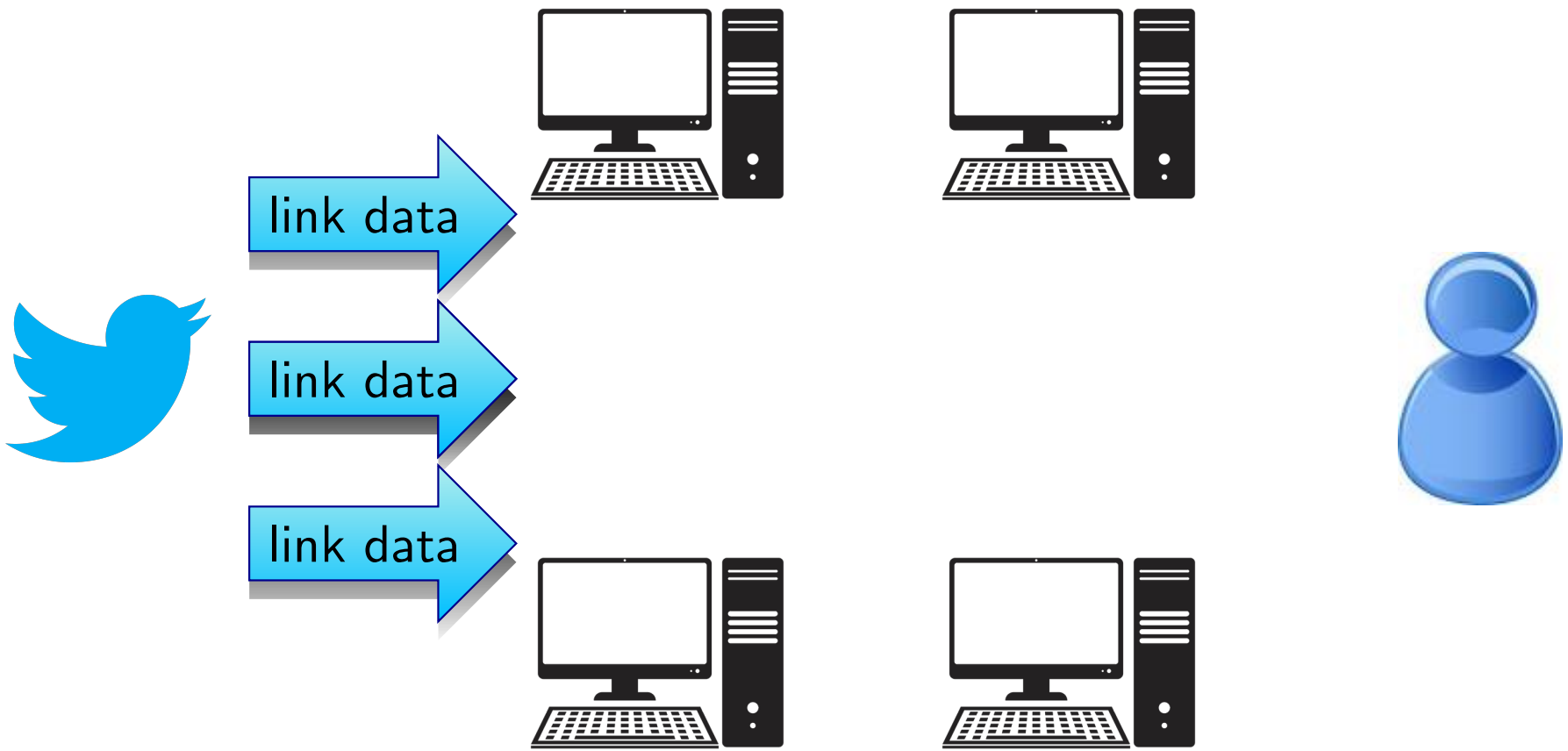


The Raphtory tool

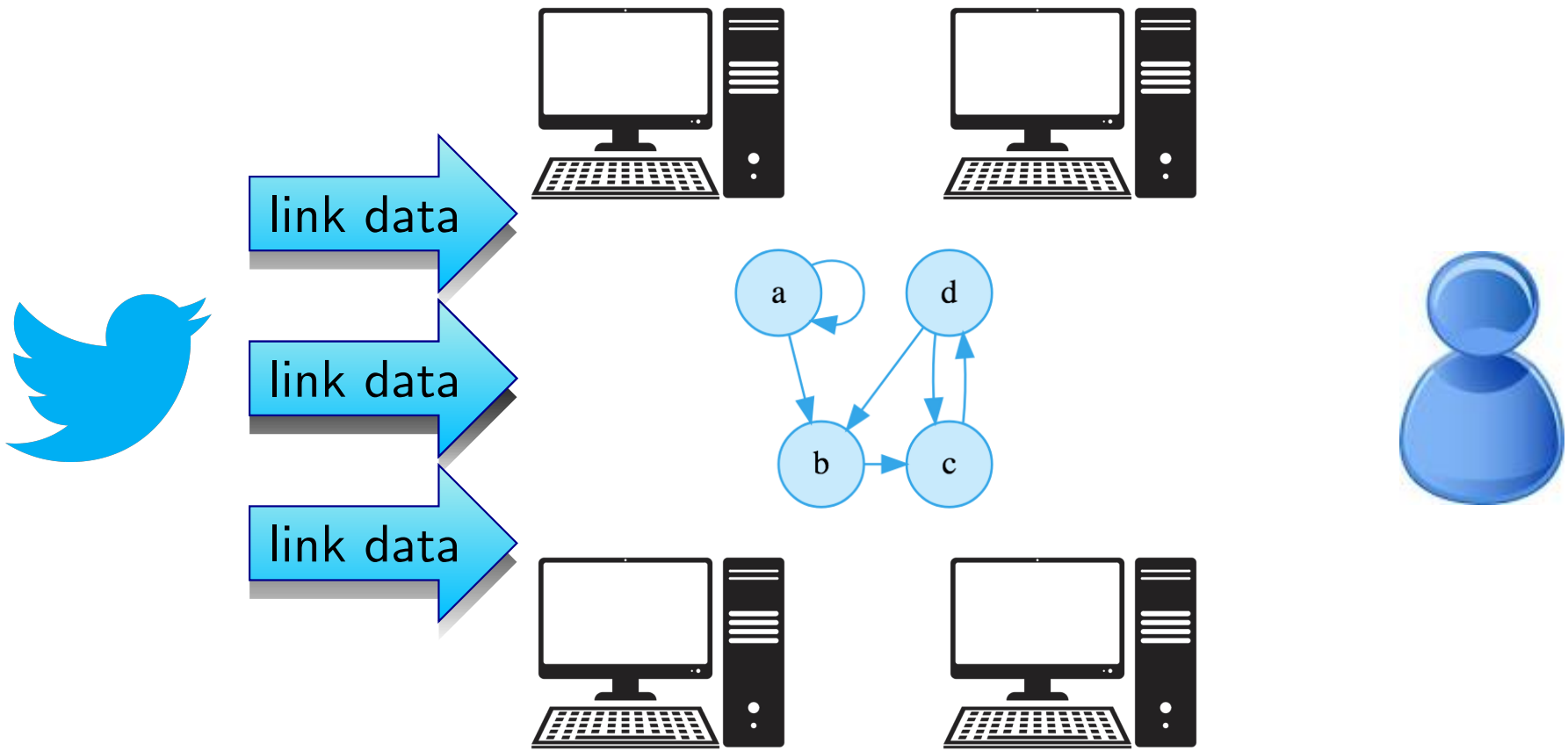


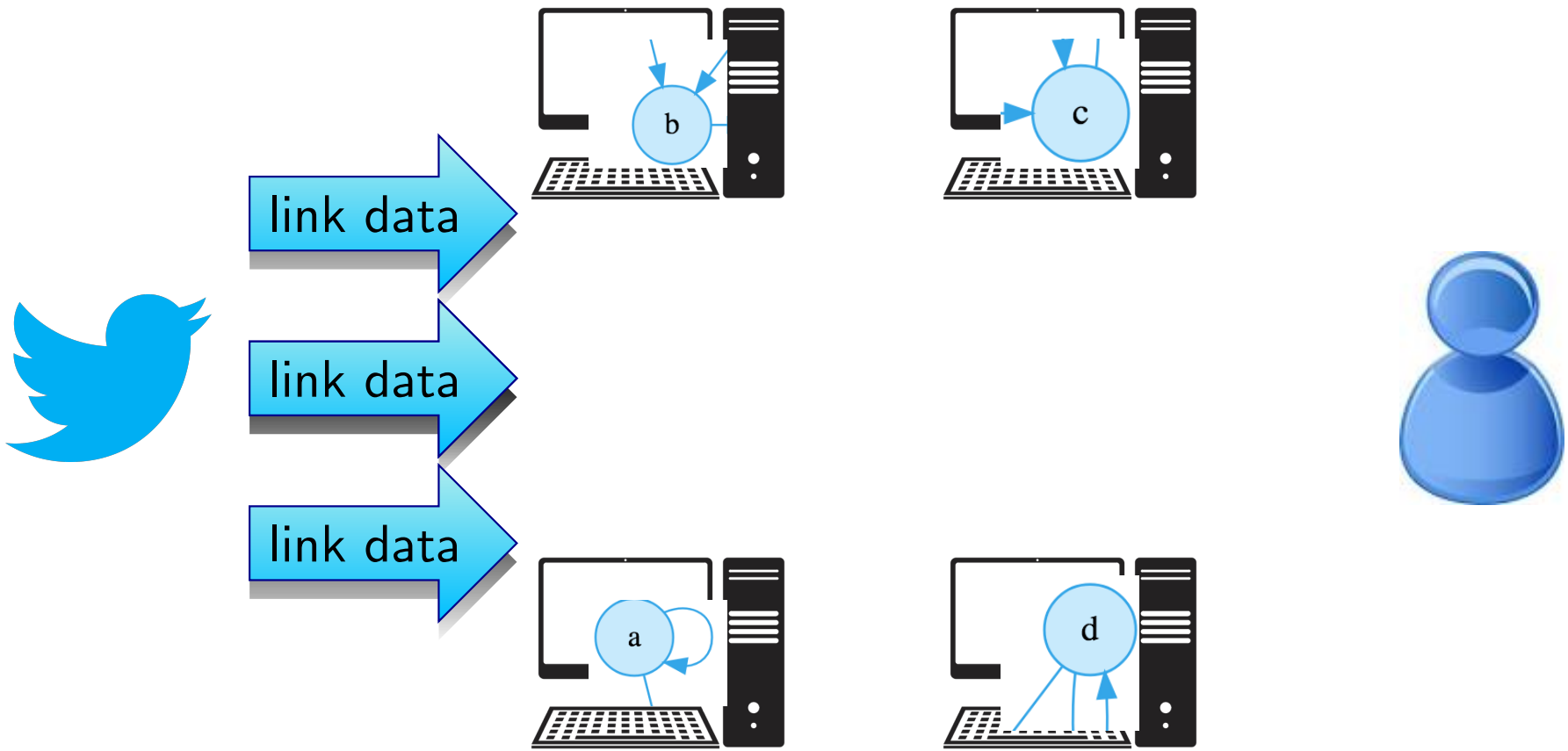
The Raphtory tool



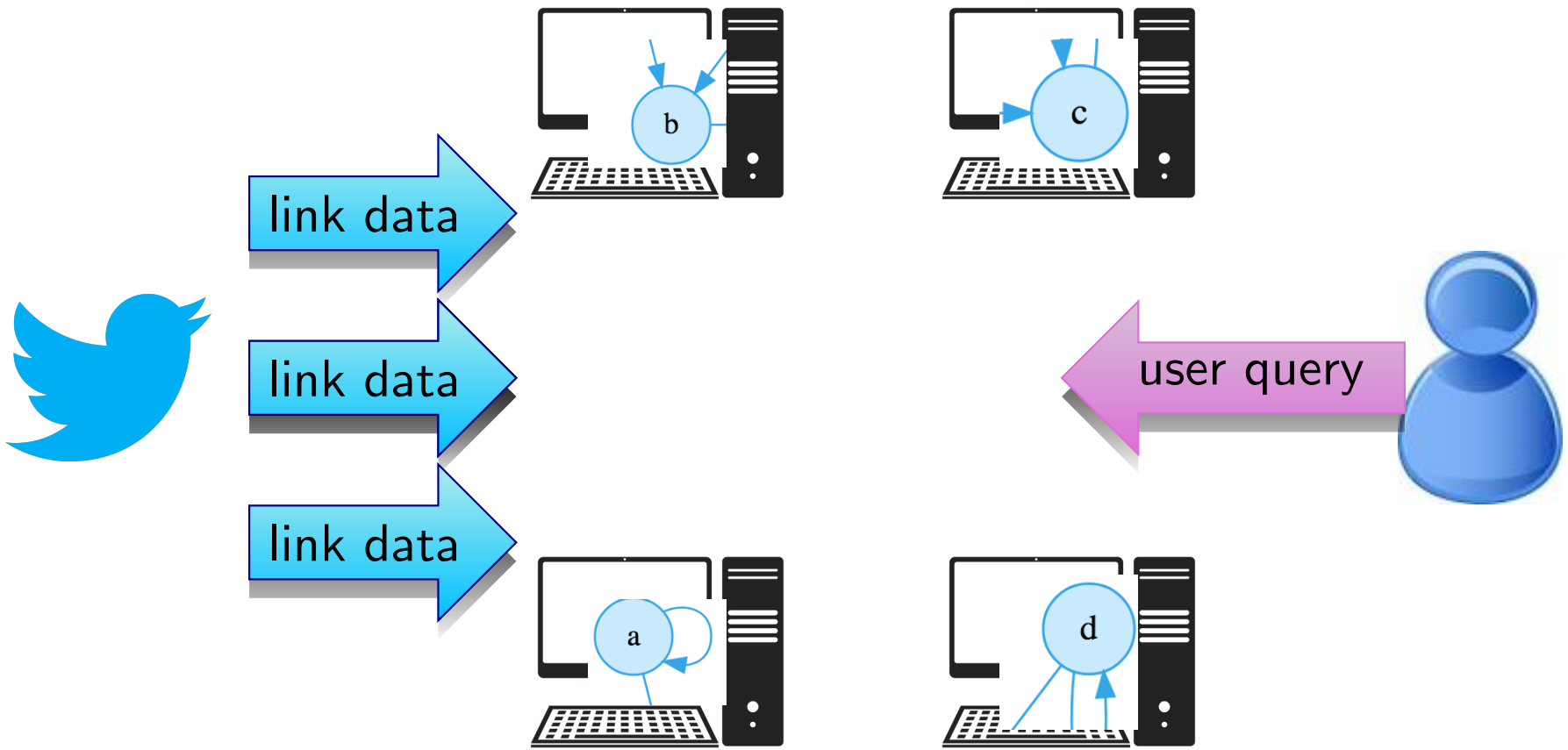
The Raphtory tool



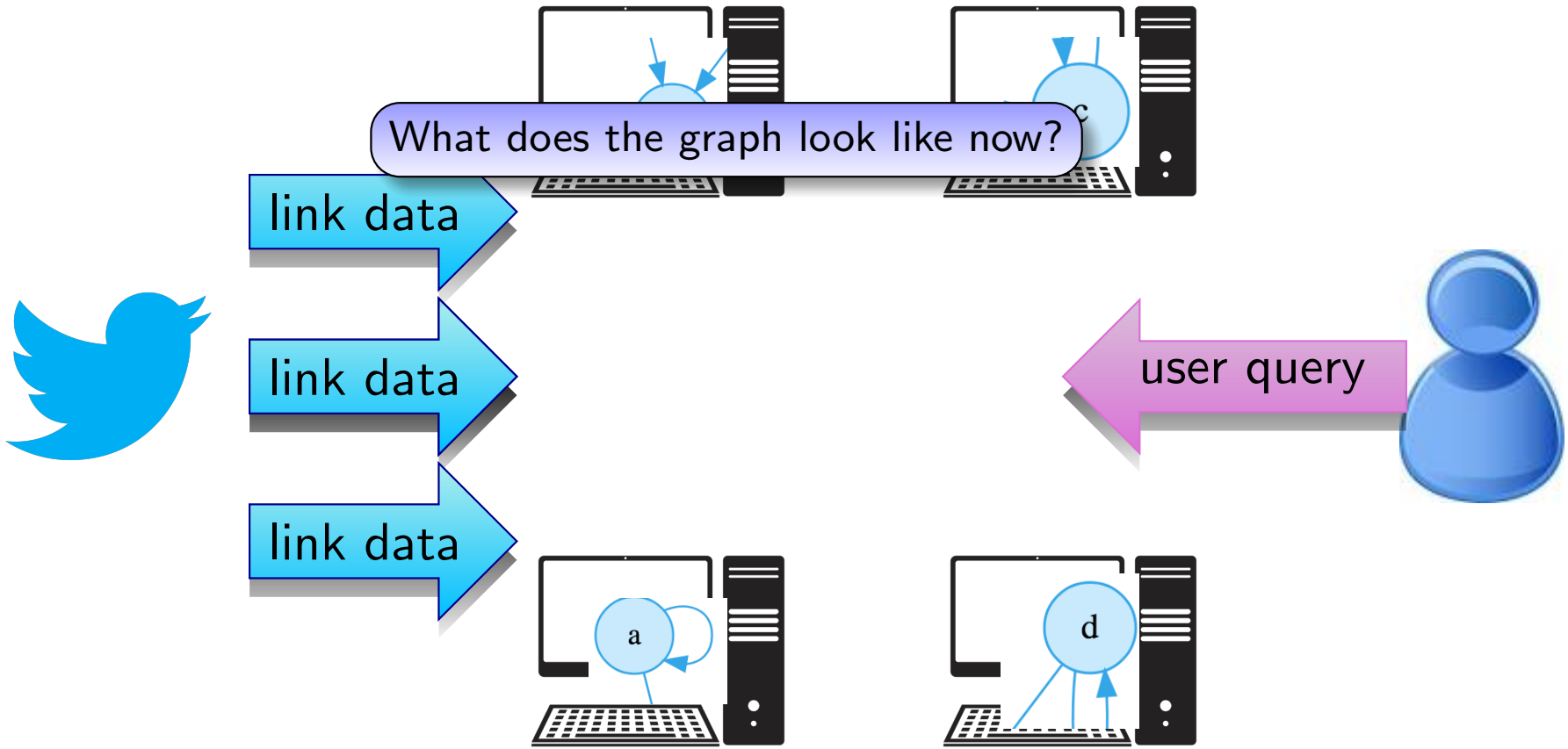
The Raphtory tool



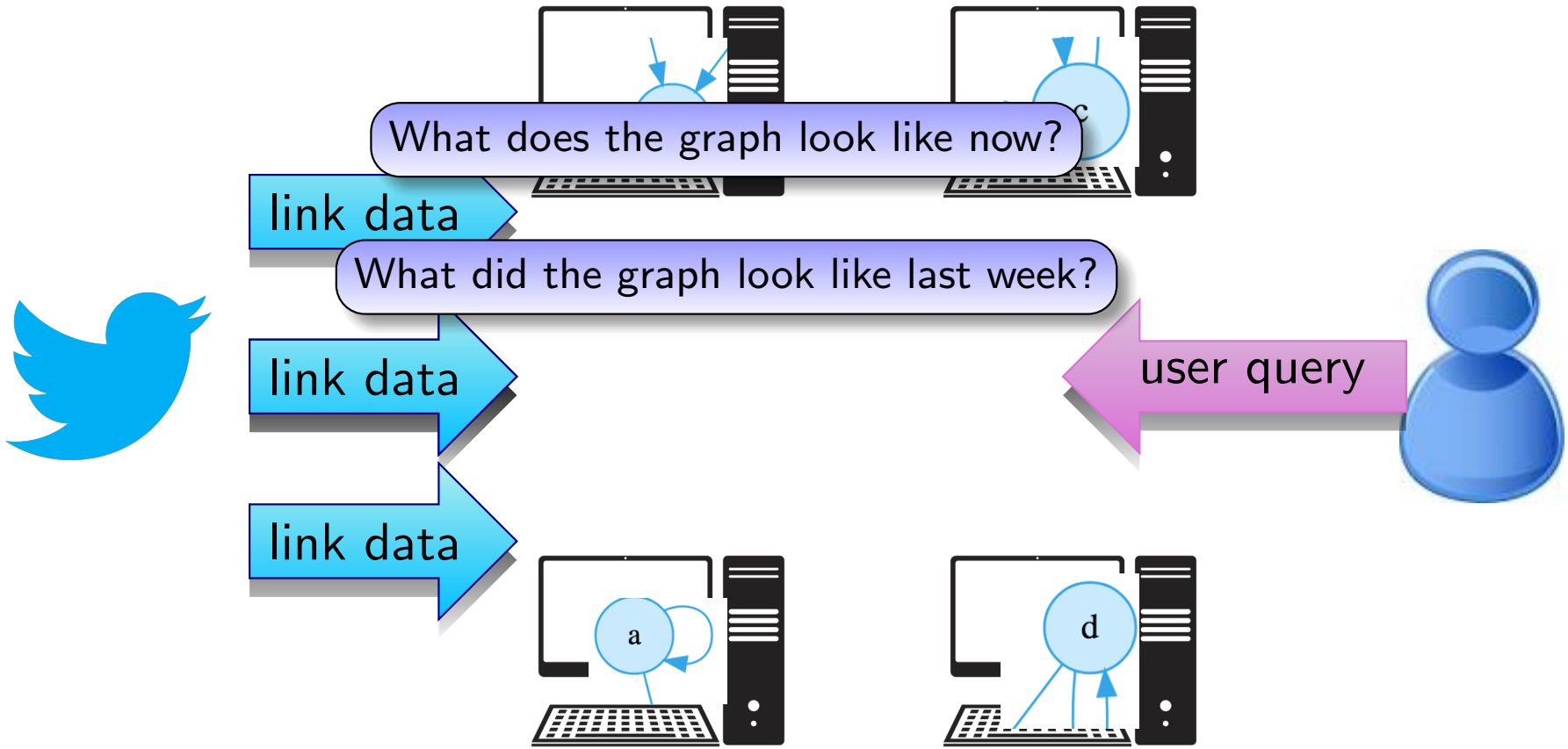
The Raphtory tool



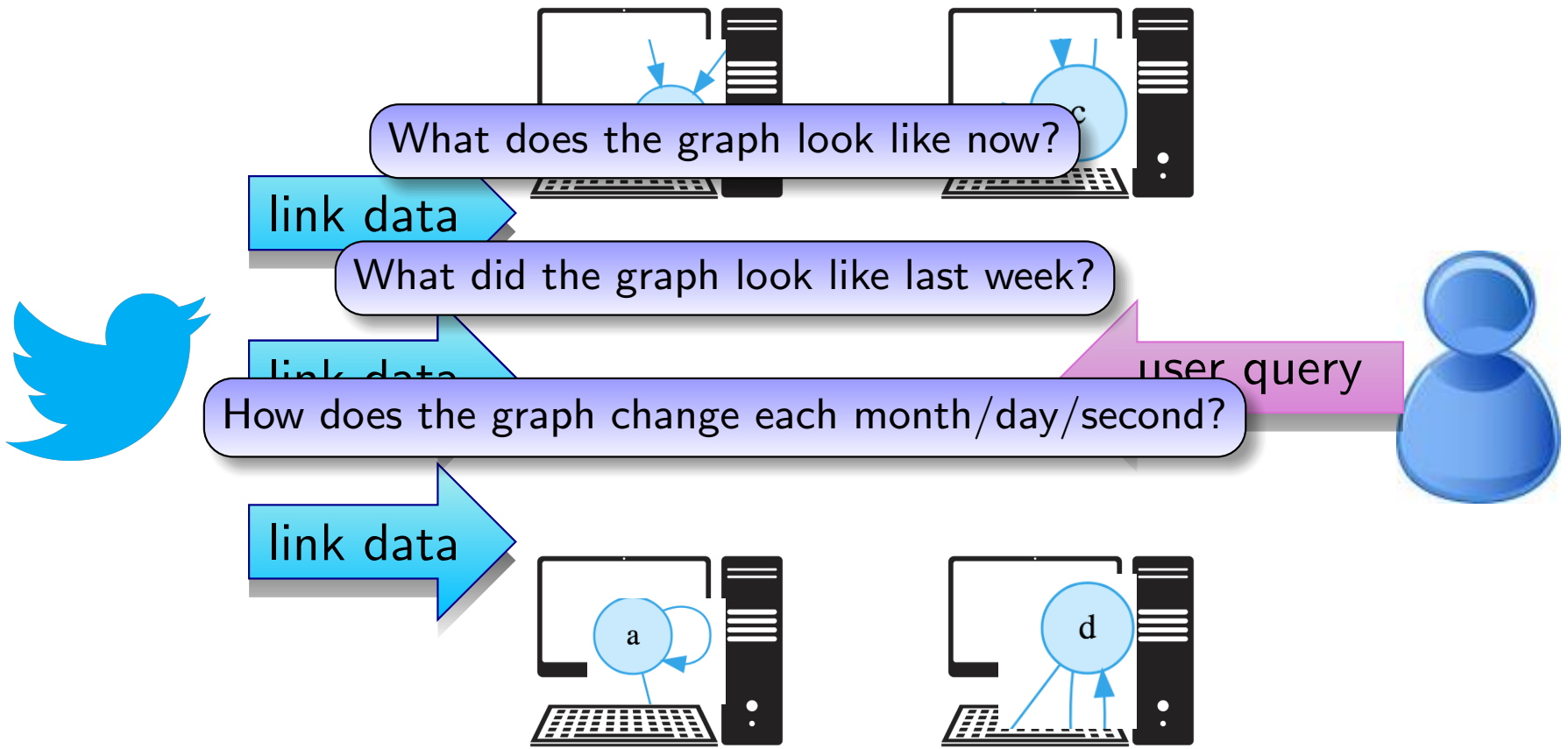
The Raphtory tool



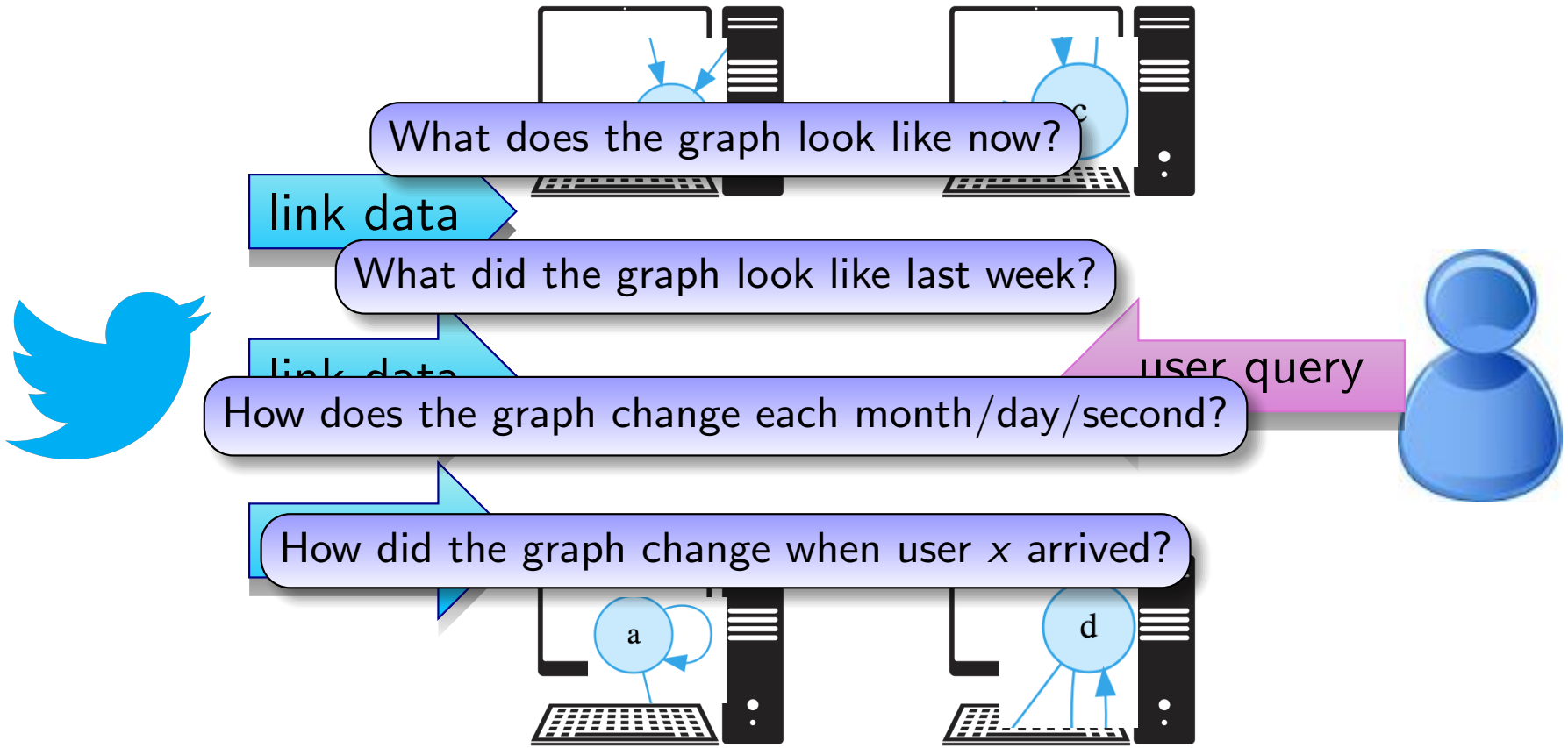
The Raphtory tool



The Raphtory tool



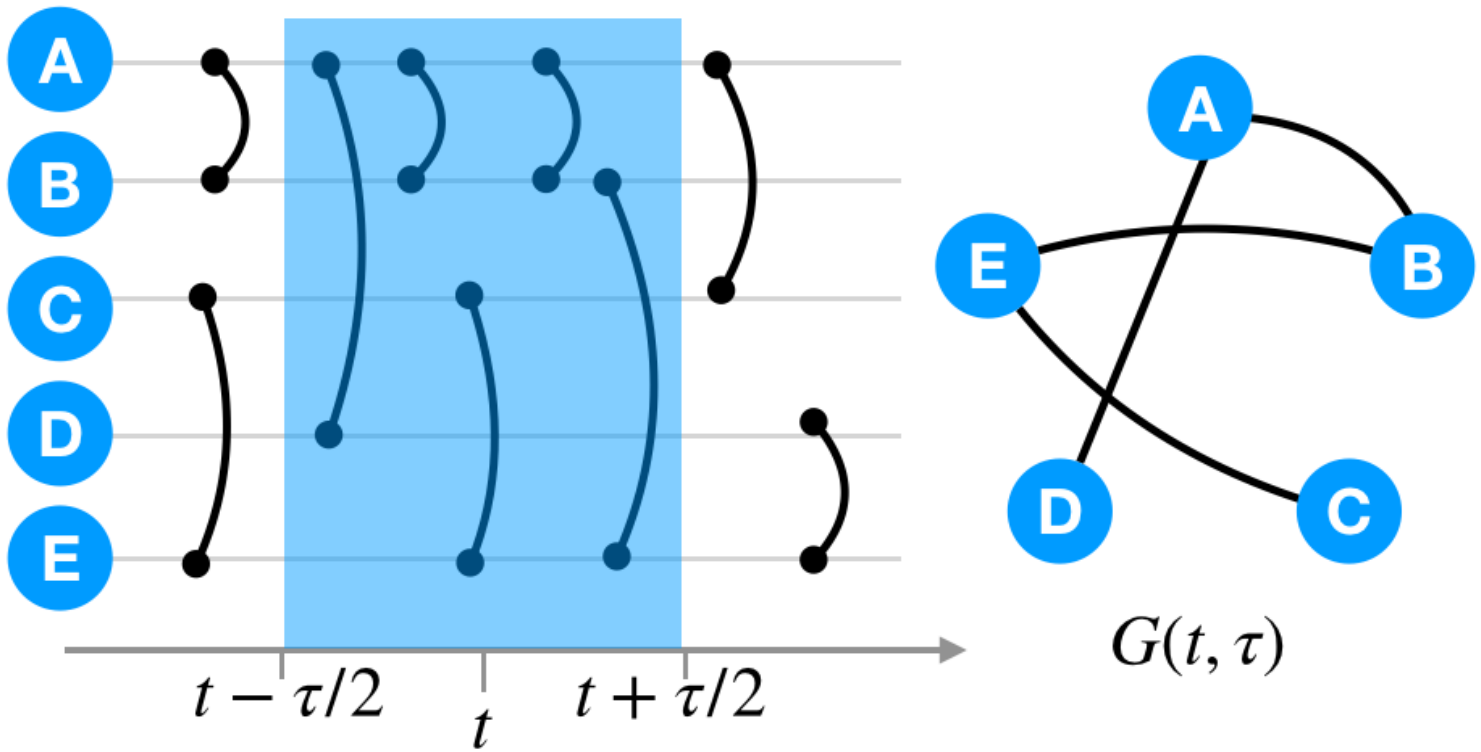
The Raphtory tool



Temporal graphs – interaction graphs

Interaction graph for time t and a window length τ

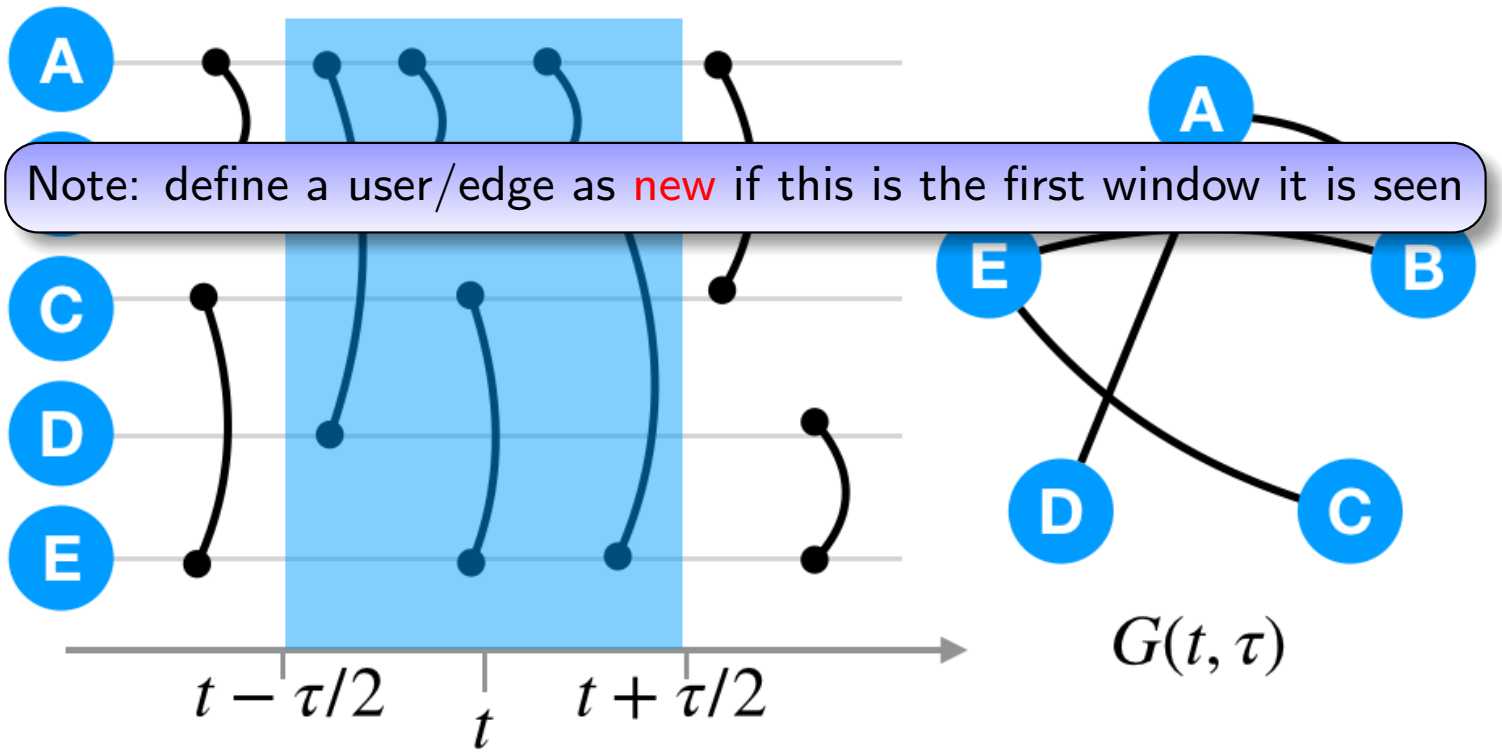
The graph $G(t, \tau)$ is defined by the set of all edges i, j where i and j interact at a time T such that $t - \tau/2 \leq T \leq t + \tau/2$.



Temporal graphs – interaction graphs

Interaction graph for time t and a window length τ

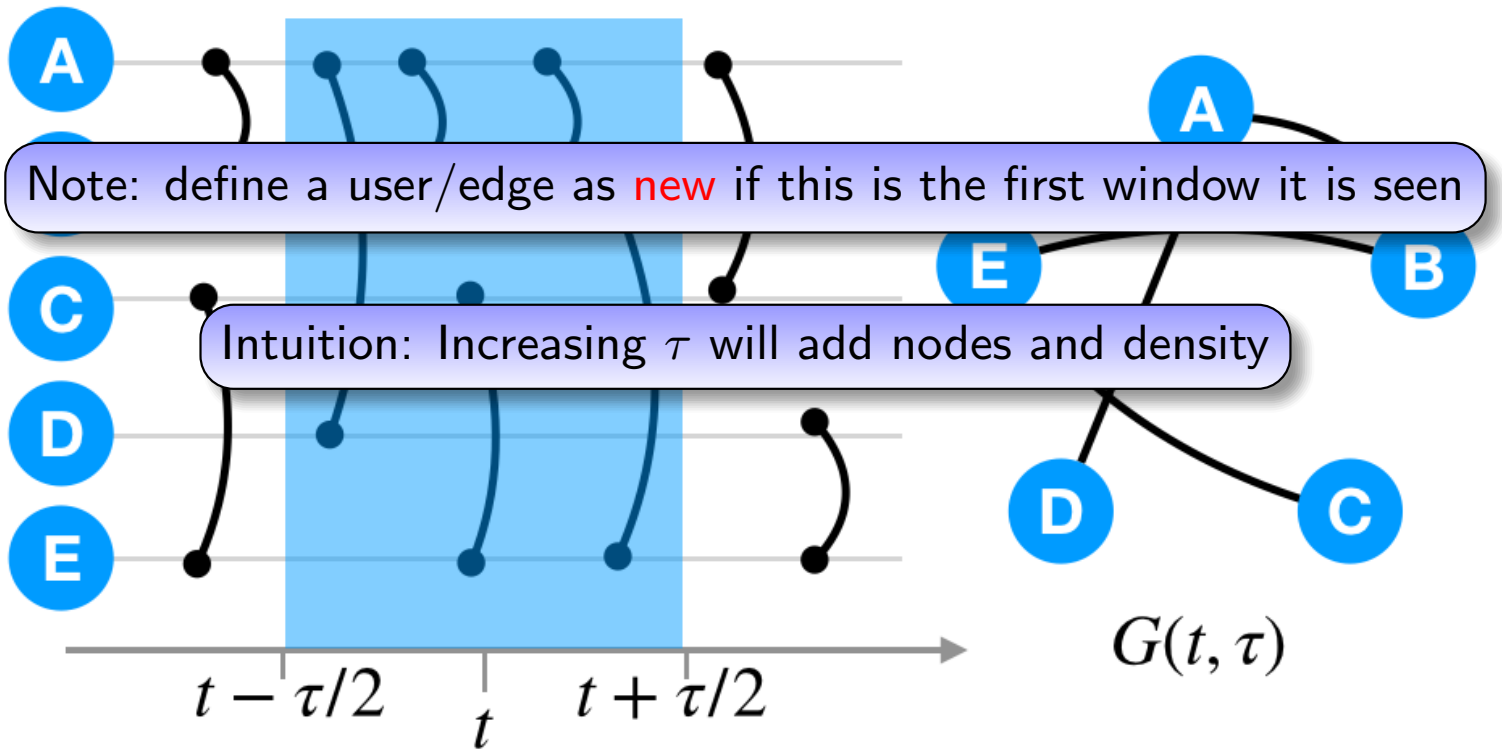
The graph $G(t, \tau)$ is defined by the set of all edges i, j where i and j interact at a time T such that $t - \tau/2 \leq T \leq t + \tau/2$.



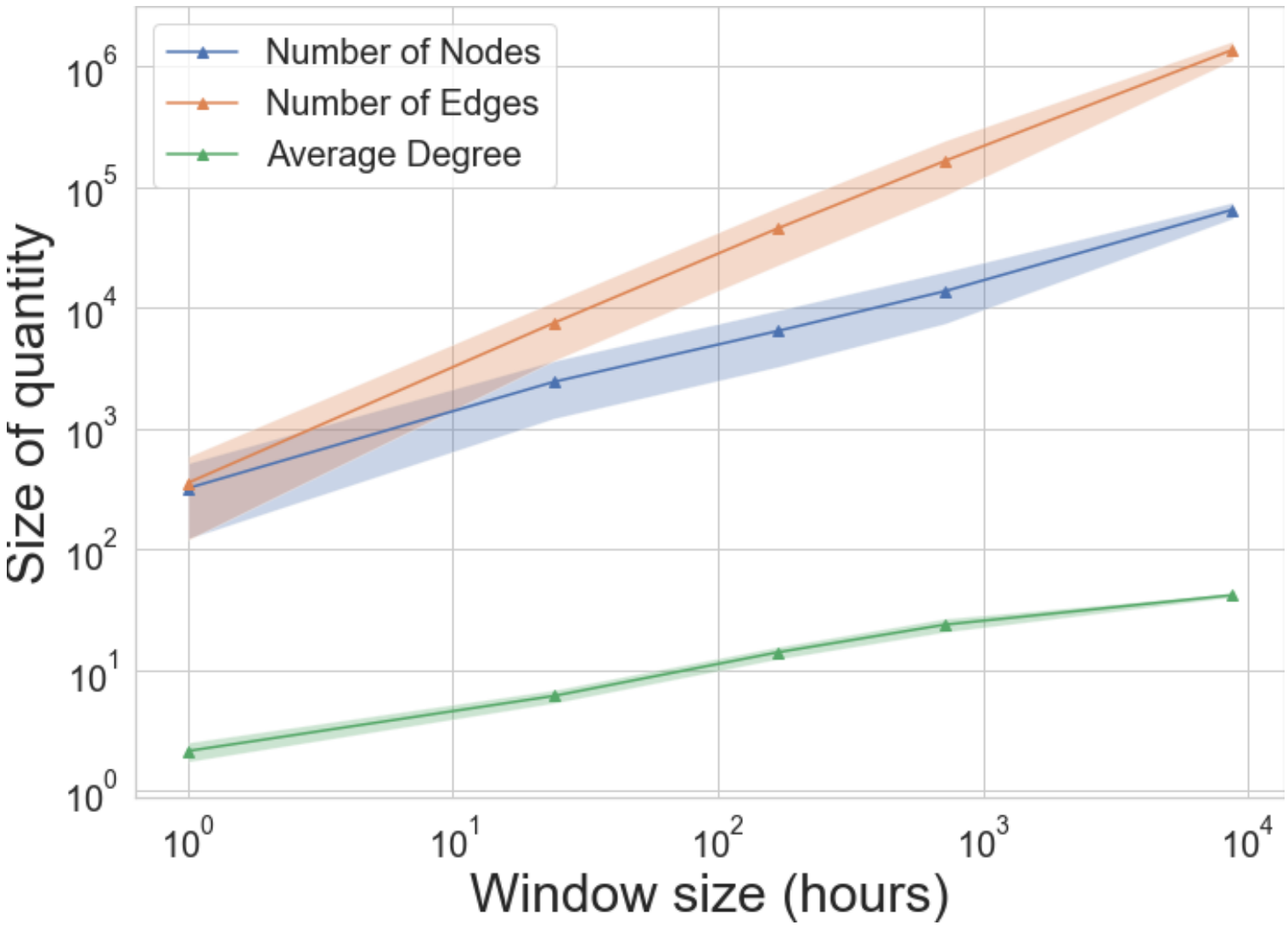
Temporal graphs – interaction graphs

Interaction graph for time t and a window length τ

The graph $G(t, \tau)$ is defined by the set of all edges i, j where i and j interact at a time T such that $t - \tau/2 \leq T \leq t + \tau/2$.



Checking our intuition



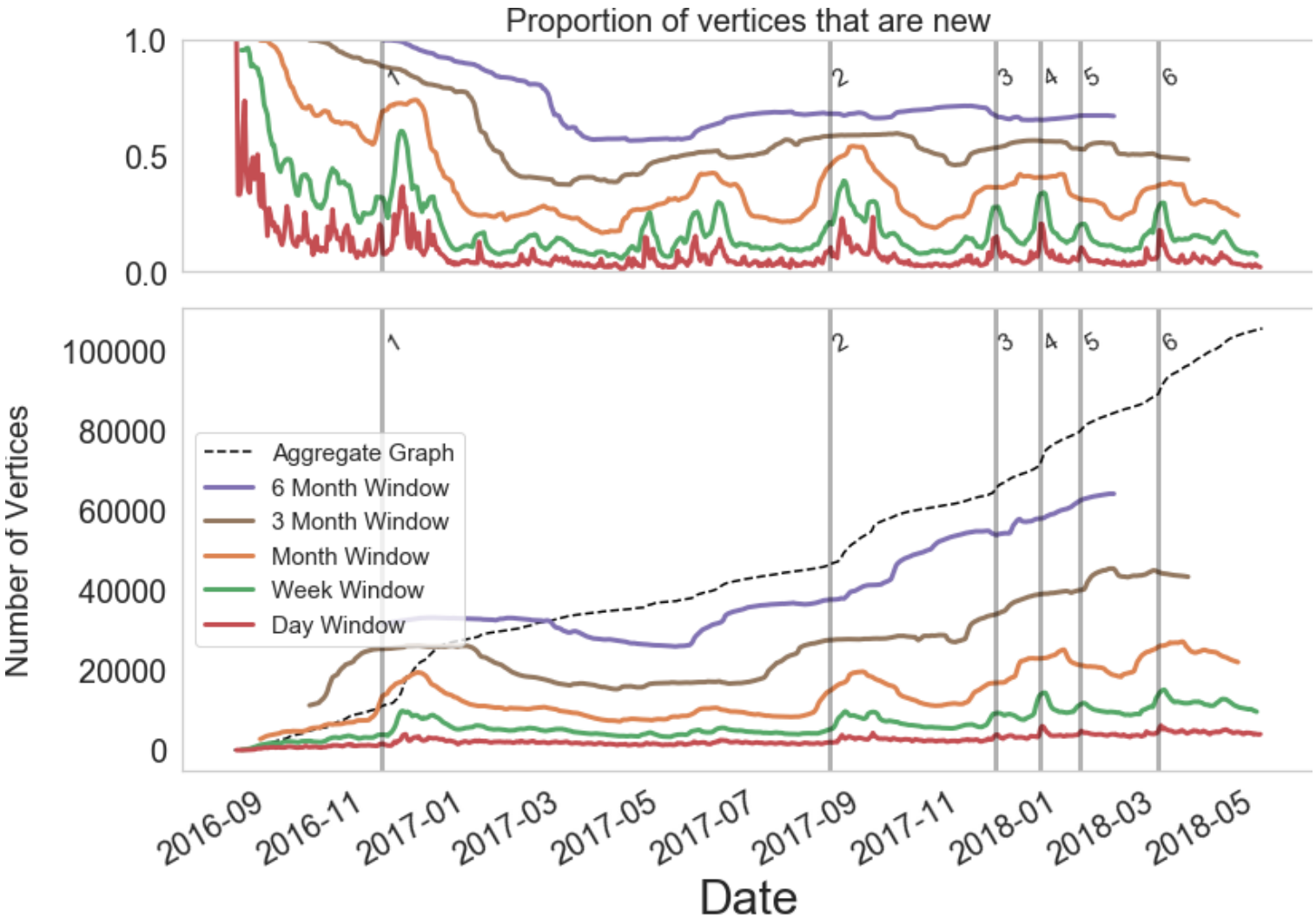
Some research questions for gab data

What simple questions can we ask of the data

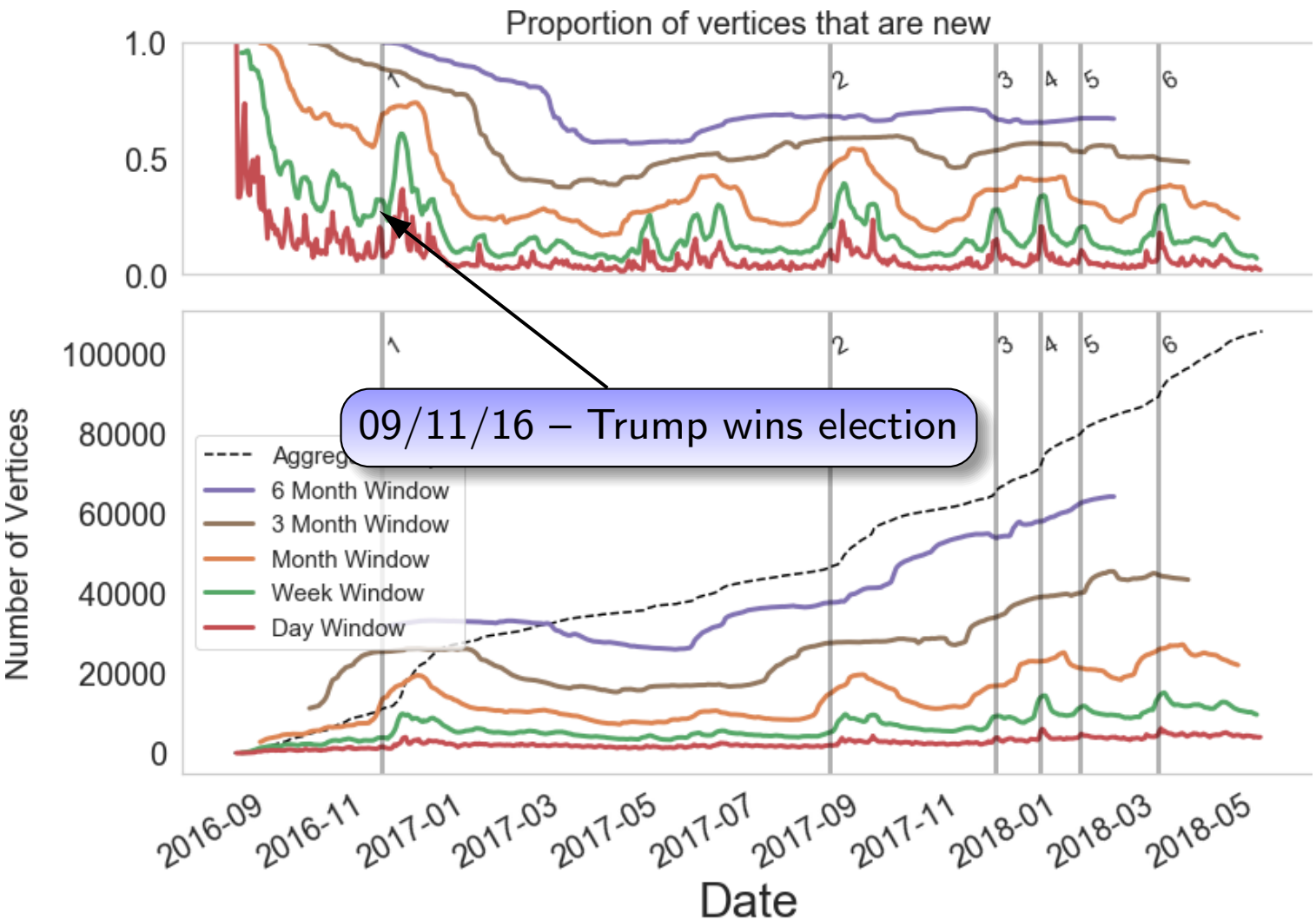
Remember our focus is structure not content. We are not (we cheat a little here at one point) digging into the content of messages.

- Is gab growing, what drives the growth?
- Is gab a "social" network (friends interacting)?
- Is gab a "community" (in a loose sense)?
- Is gab controlled by an "elite"?

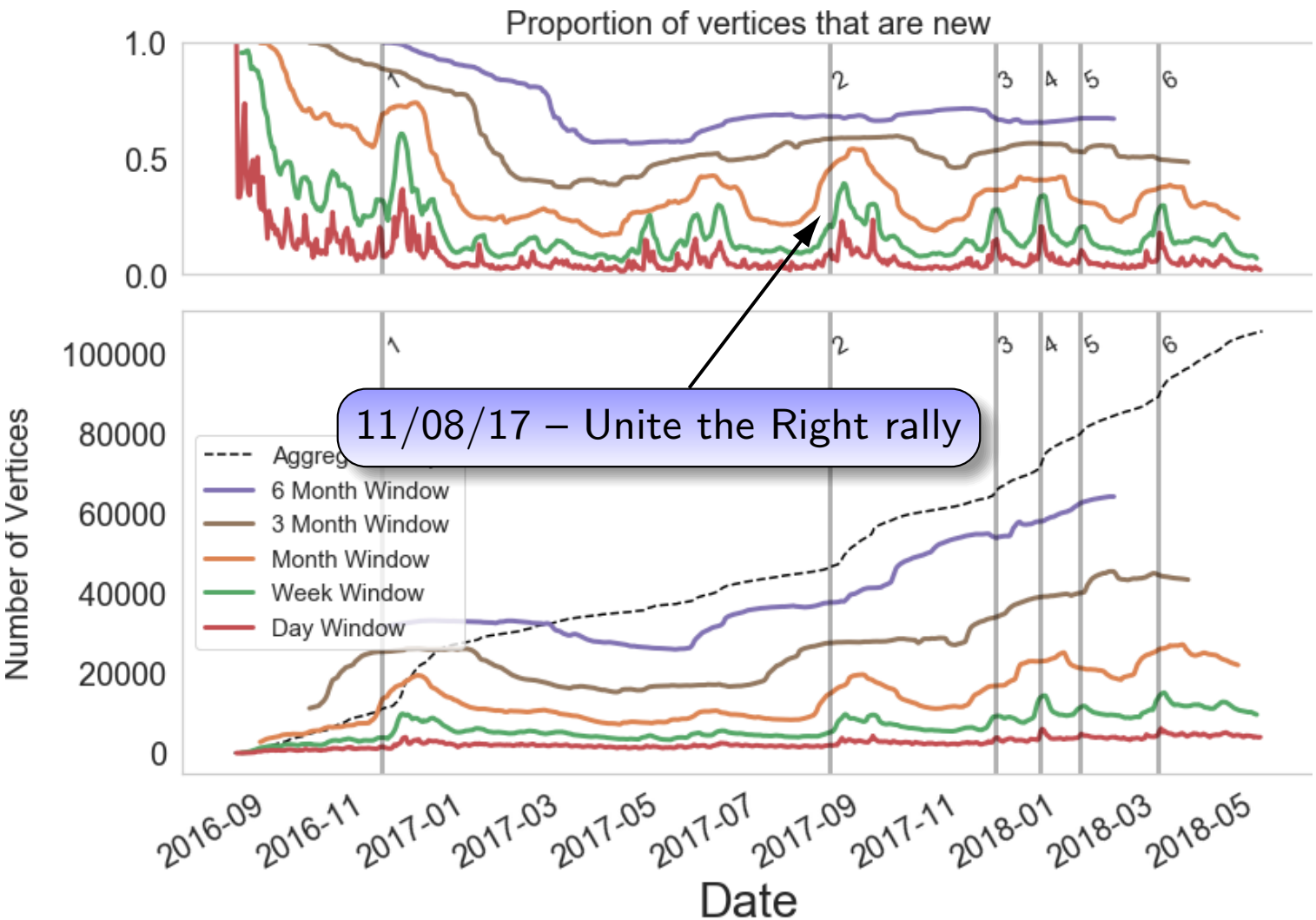
What drives growth on gab?



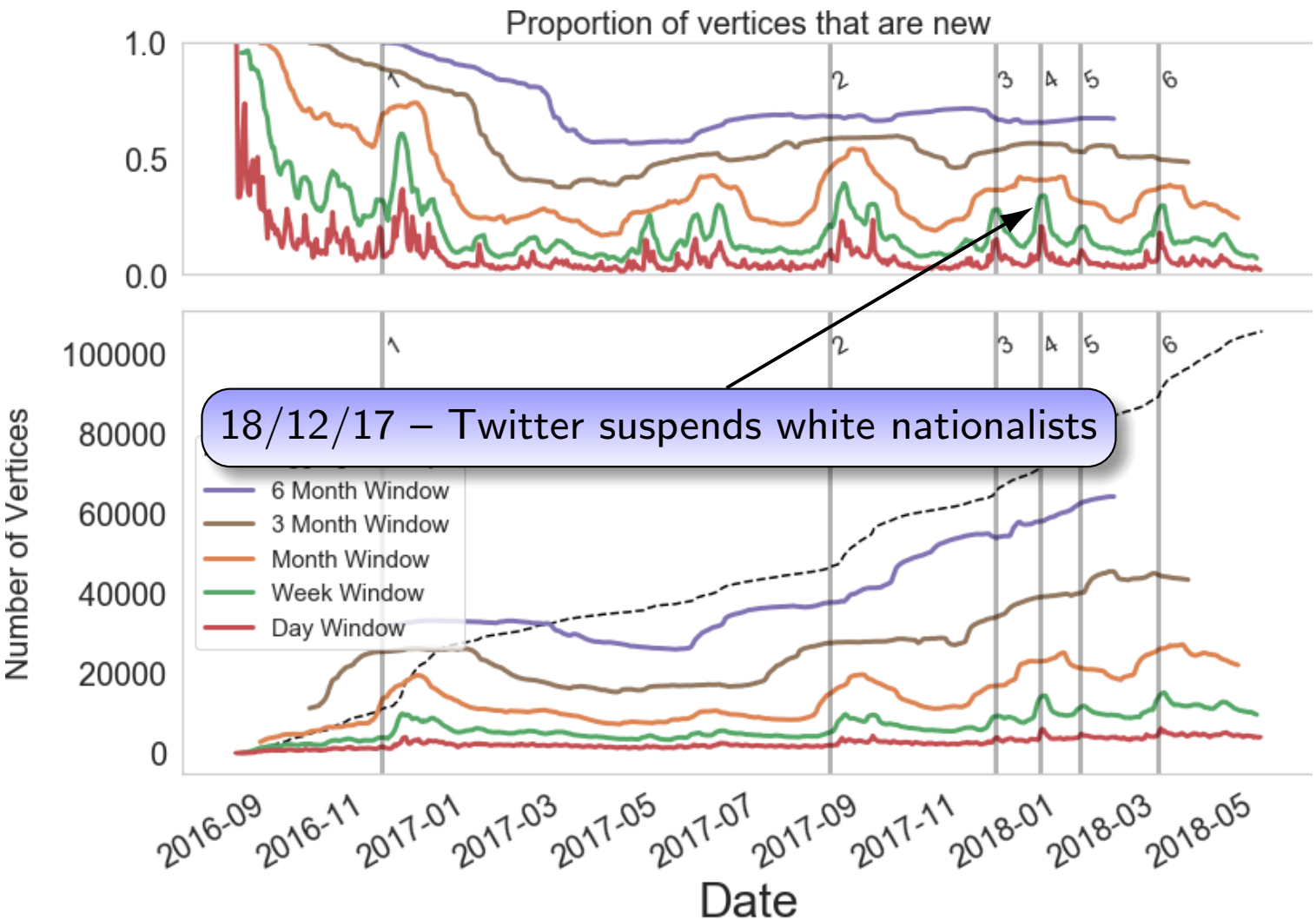
What drives growth on gab?



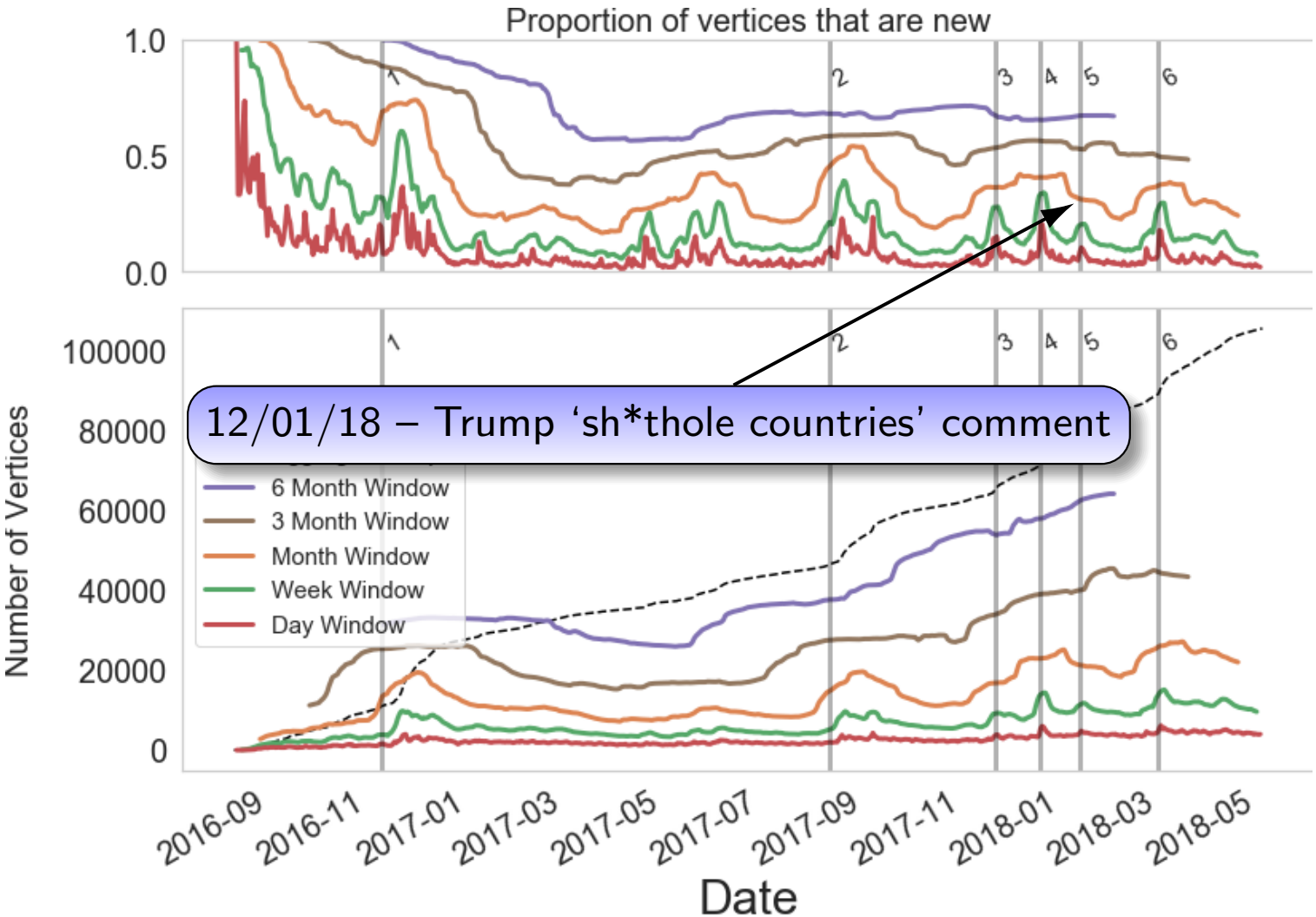
What drives growth on gab?



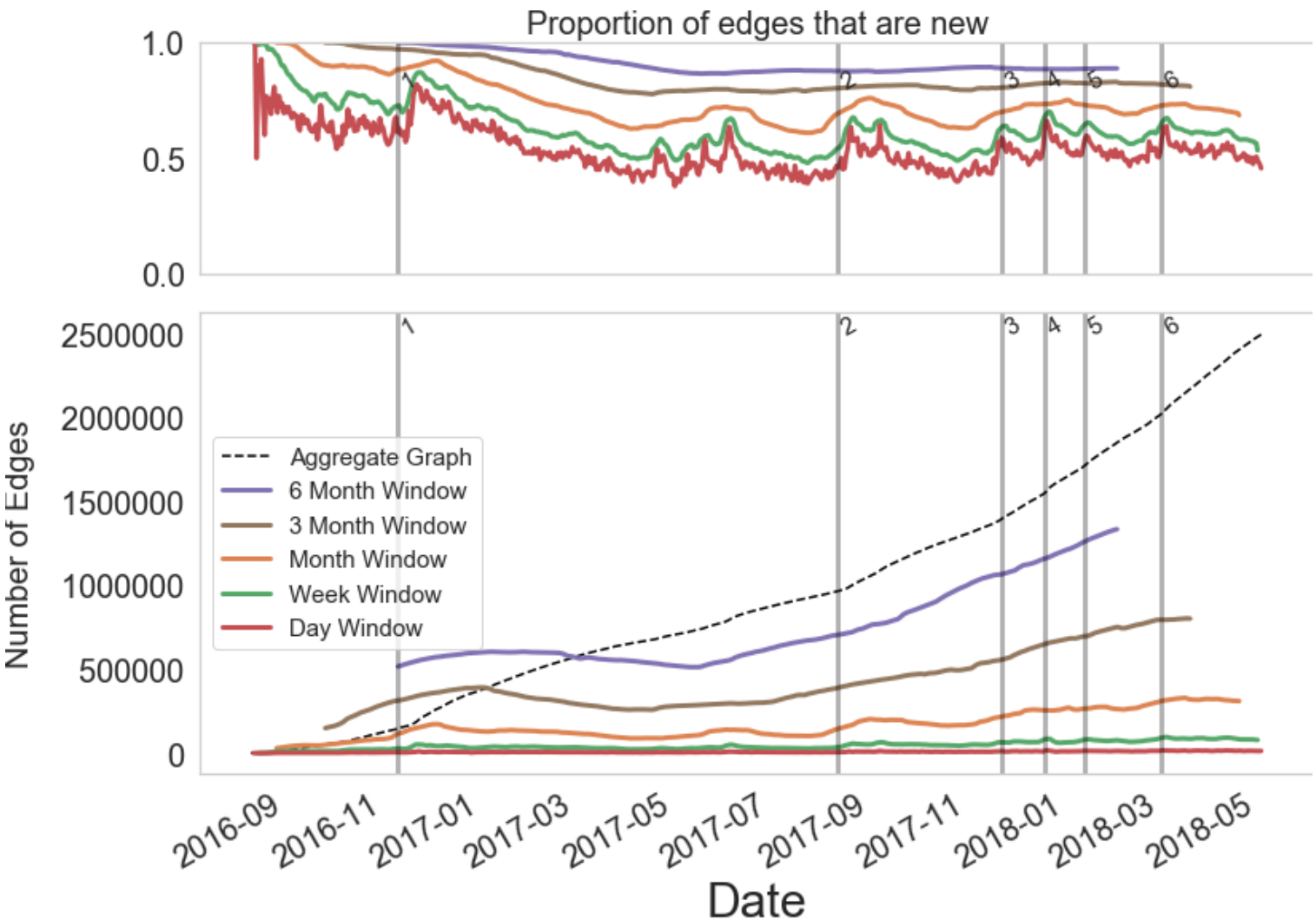
What drives growth on gab?



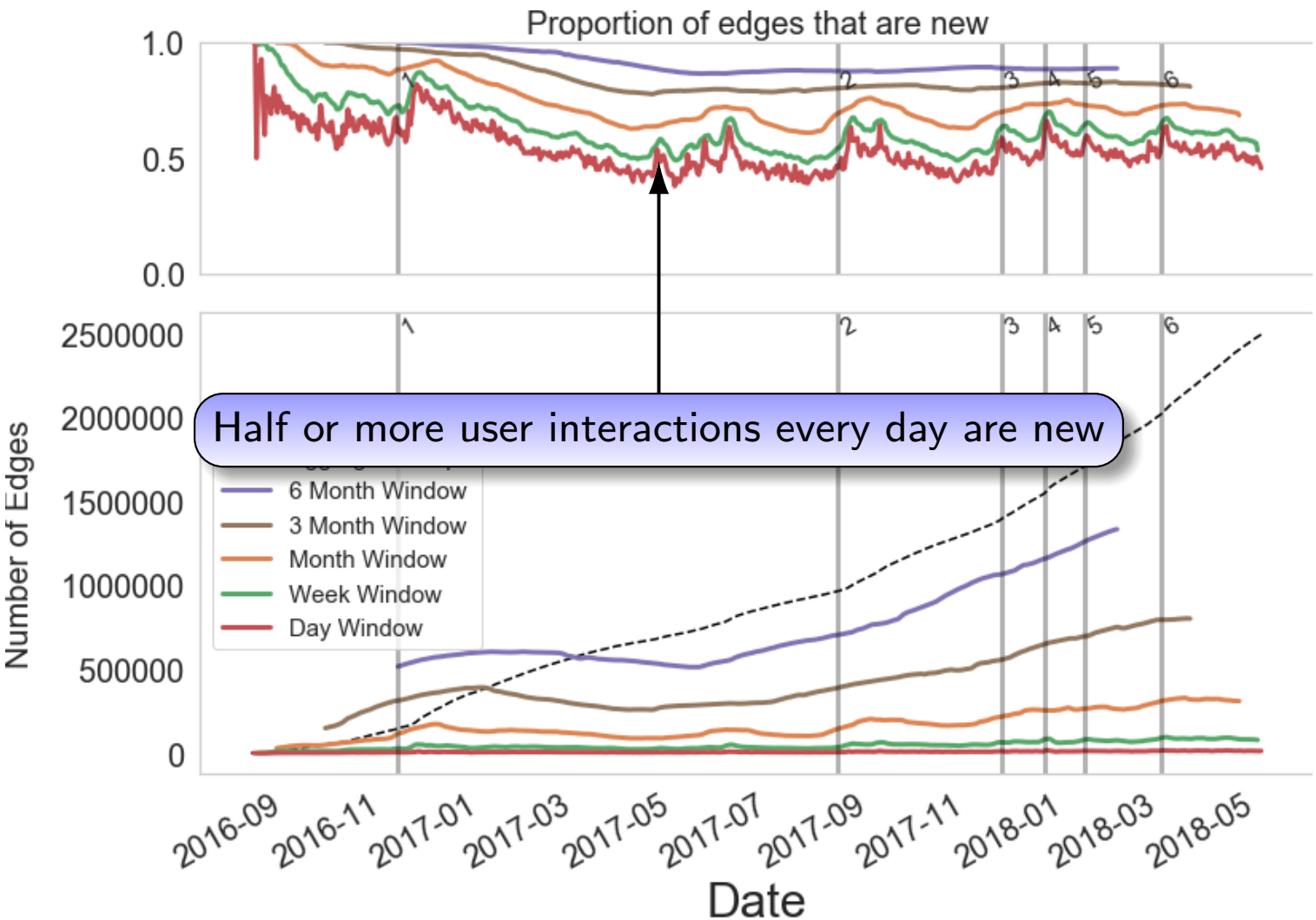
What drives growth on gab?



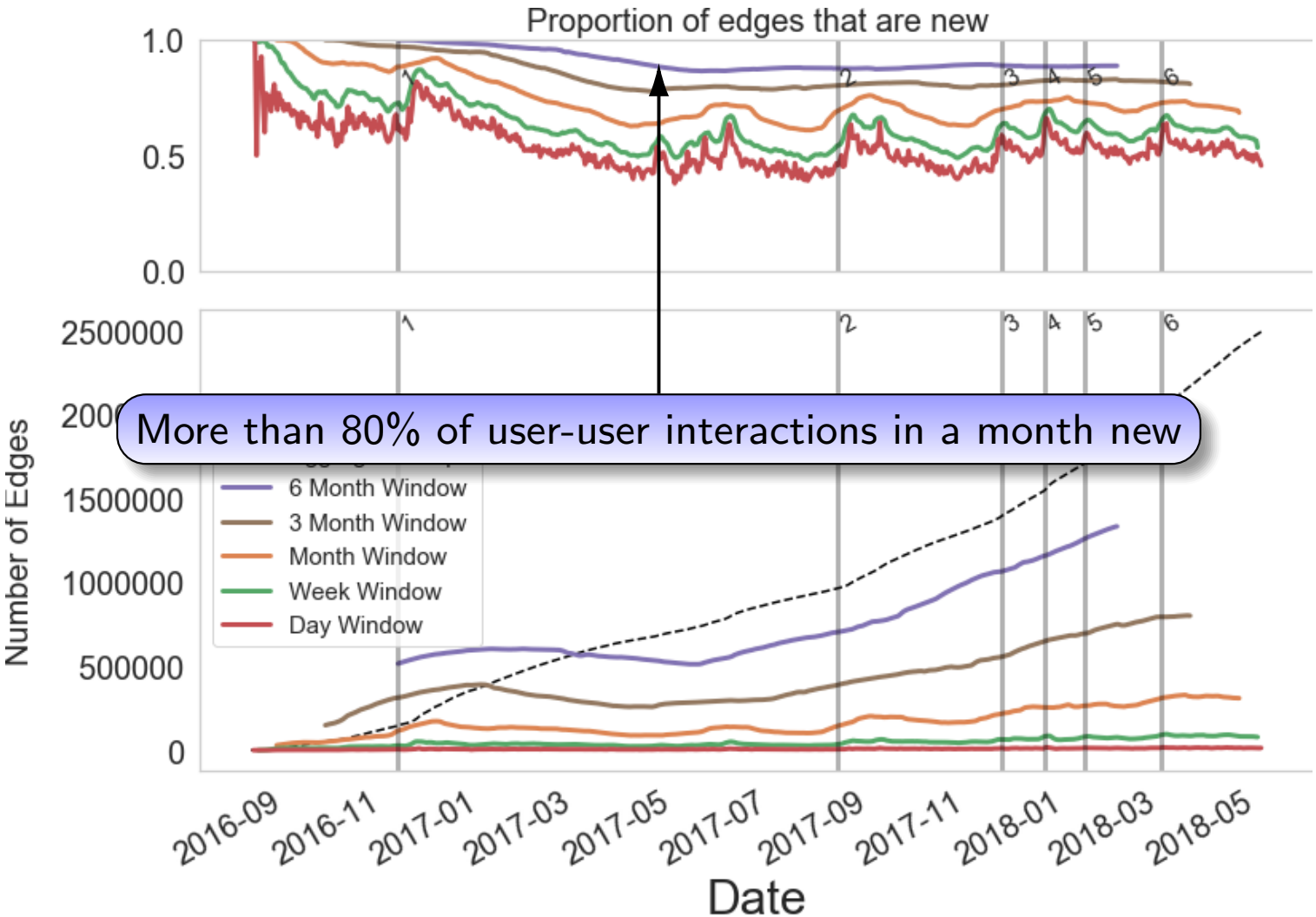
Is gab a "social" network?



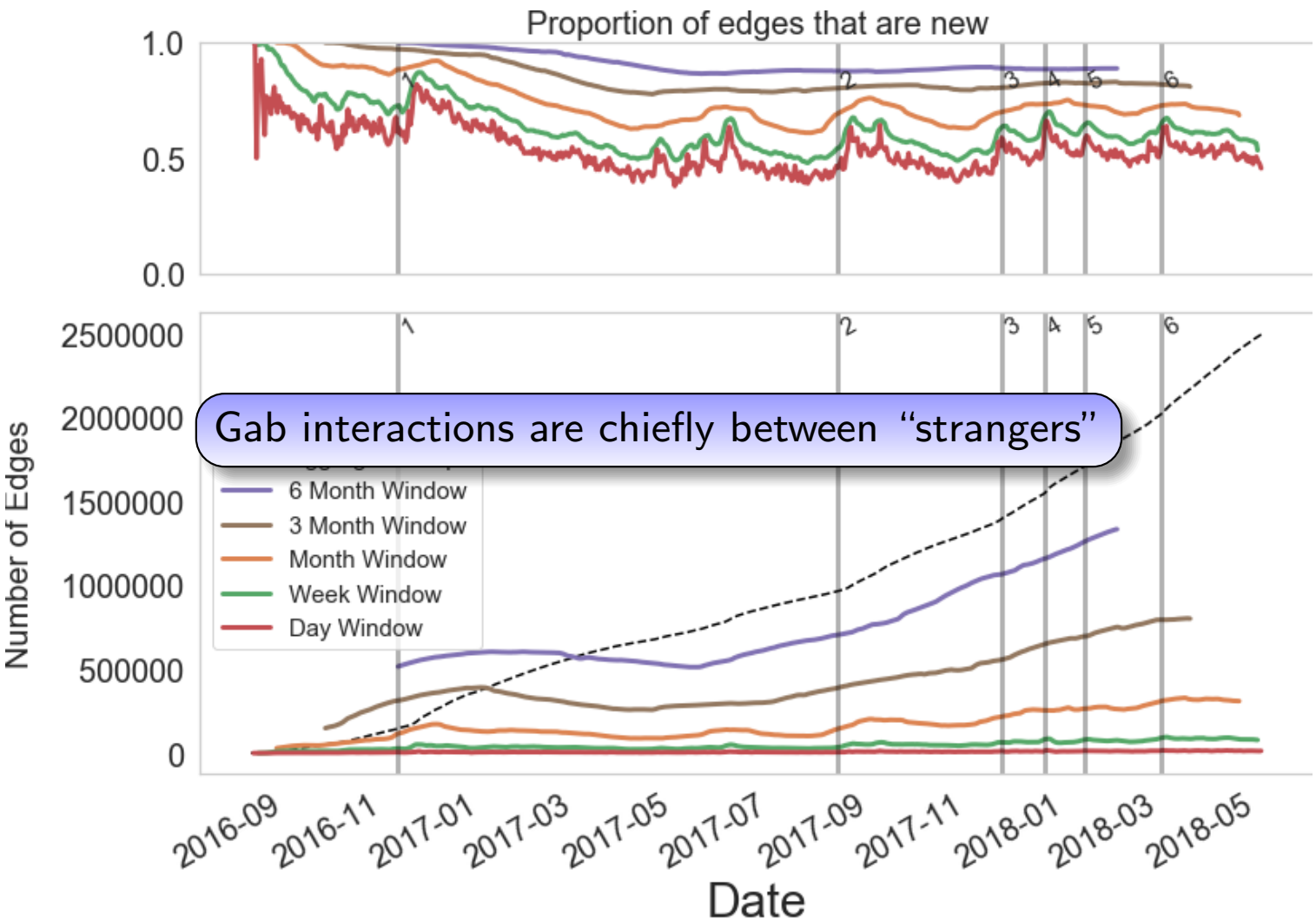
Is gab a "social" network?



Is gab a "social" network?



Is gab a "social" network?



Is gab a community? Largest Connected Component

Definition:Connected component

A connected component is a sub graph of a graph within which all nodes can trace a path to each other. The largest connected component (LCC) is the one with the most nodes.

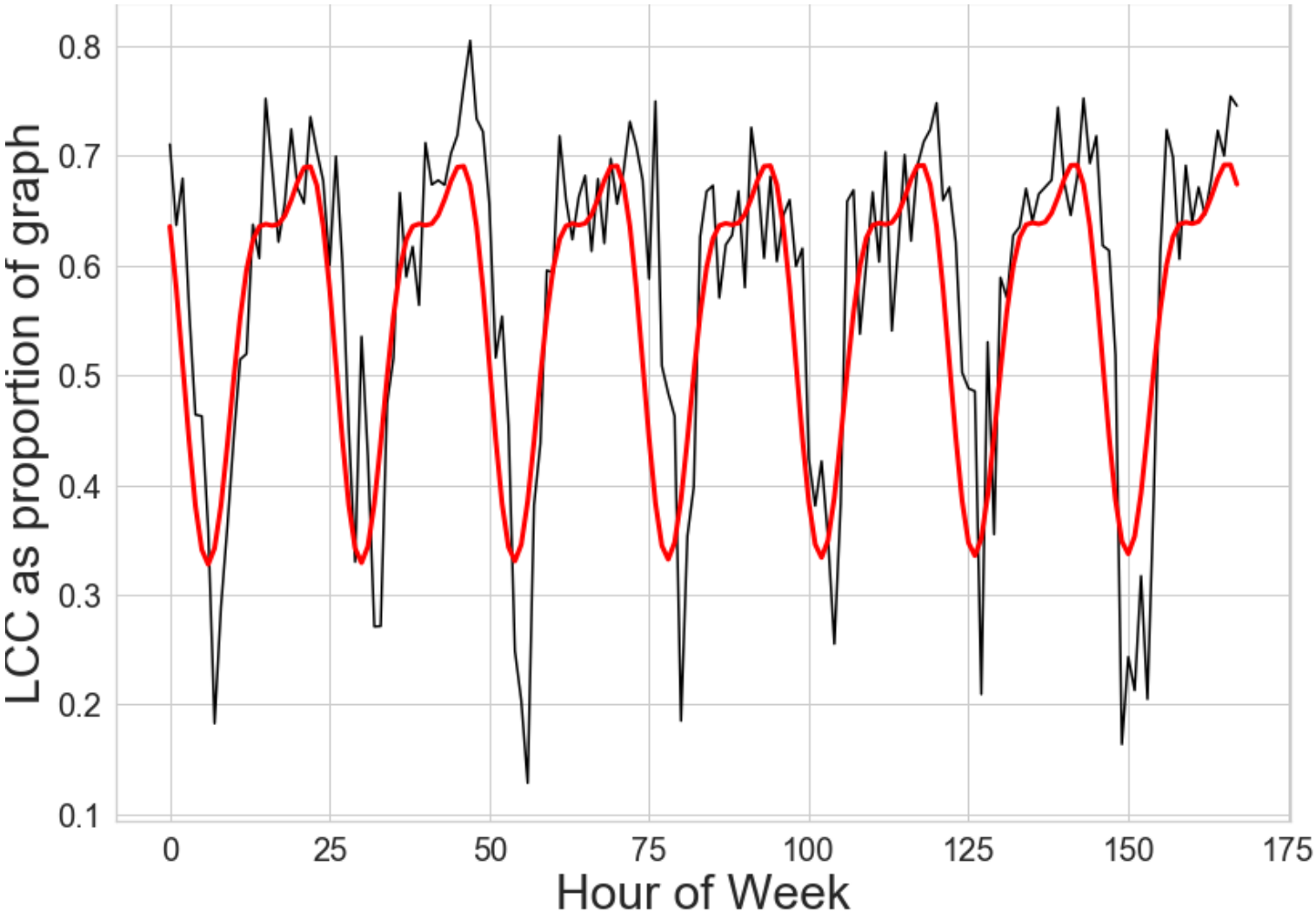
Is gab a community? Largest Connected Component

Definition:Connected component

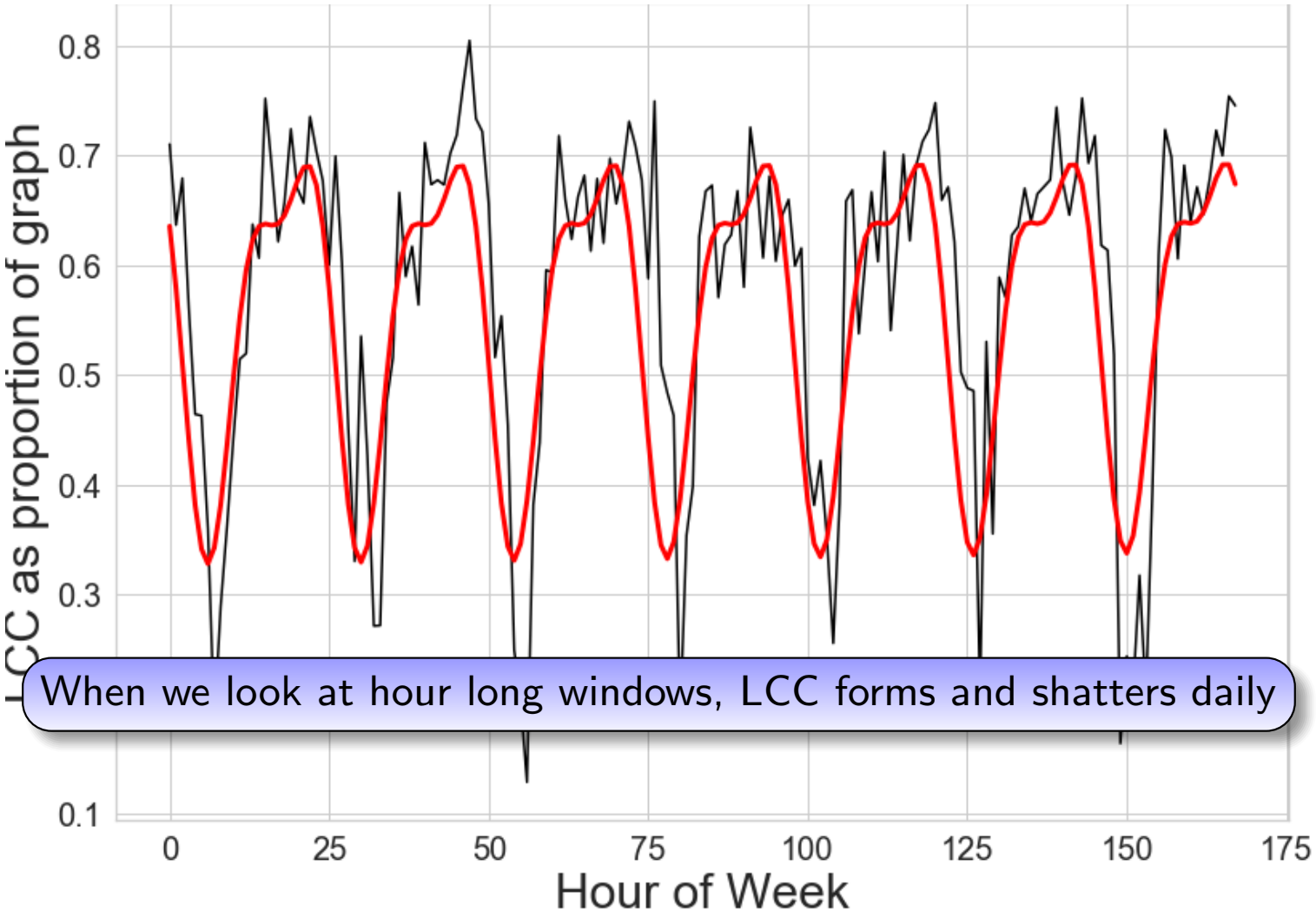
A connected component is a sub graph of a graph within which all nodes can trace a path to each other. The largest connected component (LCC) is the one with the most nodes.

- LCC is size of the largest "community" (in loosest sense).
- Remember: only count people active within the window.
- Expectation: in a "large" window most users within the LCC.
- But what happens as we look at smaller and smaller windows?

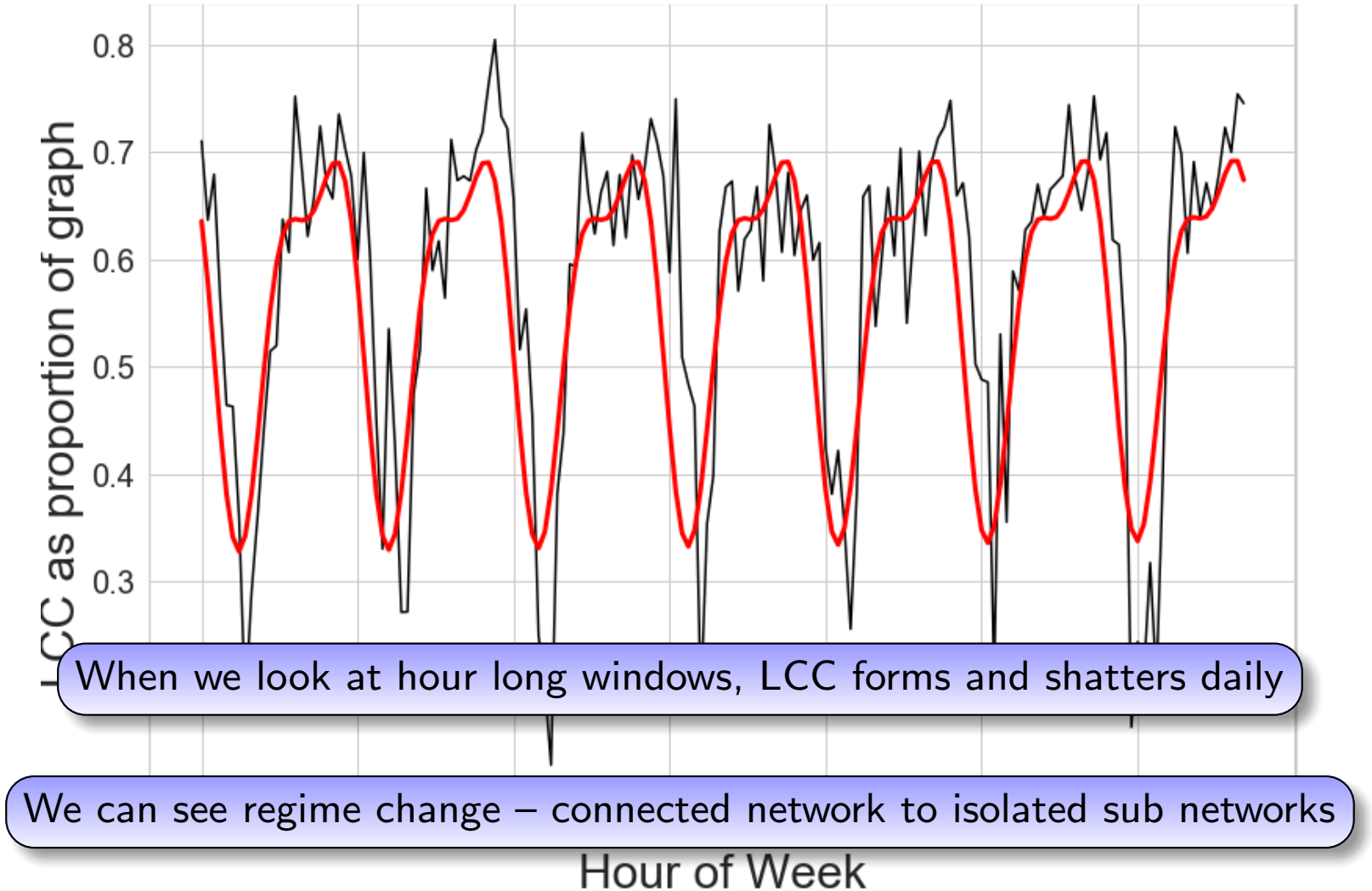
Is gab a community? LCC (hour window)



Is gab a community? LCC (hour window)



Is gab a community? LCC (hour window)



Is gab controlled by an "elite"?

- One measurement: Are the same people always most "talked about"?
- Proxy measurement: Node degree is the number of users who engaged with that user.
- Are highest degree nodes same between windows W_A and W_B?

Is gab controlled by an "elite"?

- One measurement: Are the same people always most "talked about"?
- Proxy measurement: Node degree is the number of users who engaged with that user.
- Are highest degree nodes same between windows W_A and W_B?

Jaccard Similarity (for top N users)

Let A, B be set of top N users in windows W_A, W_B.

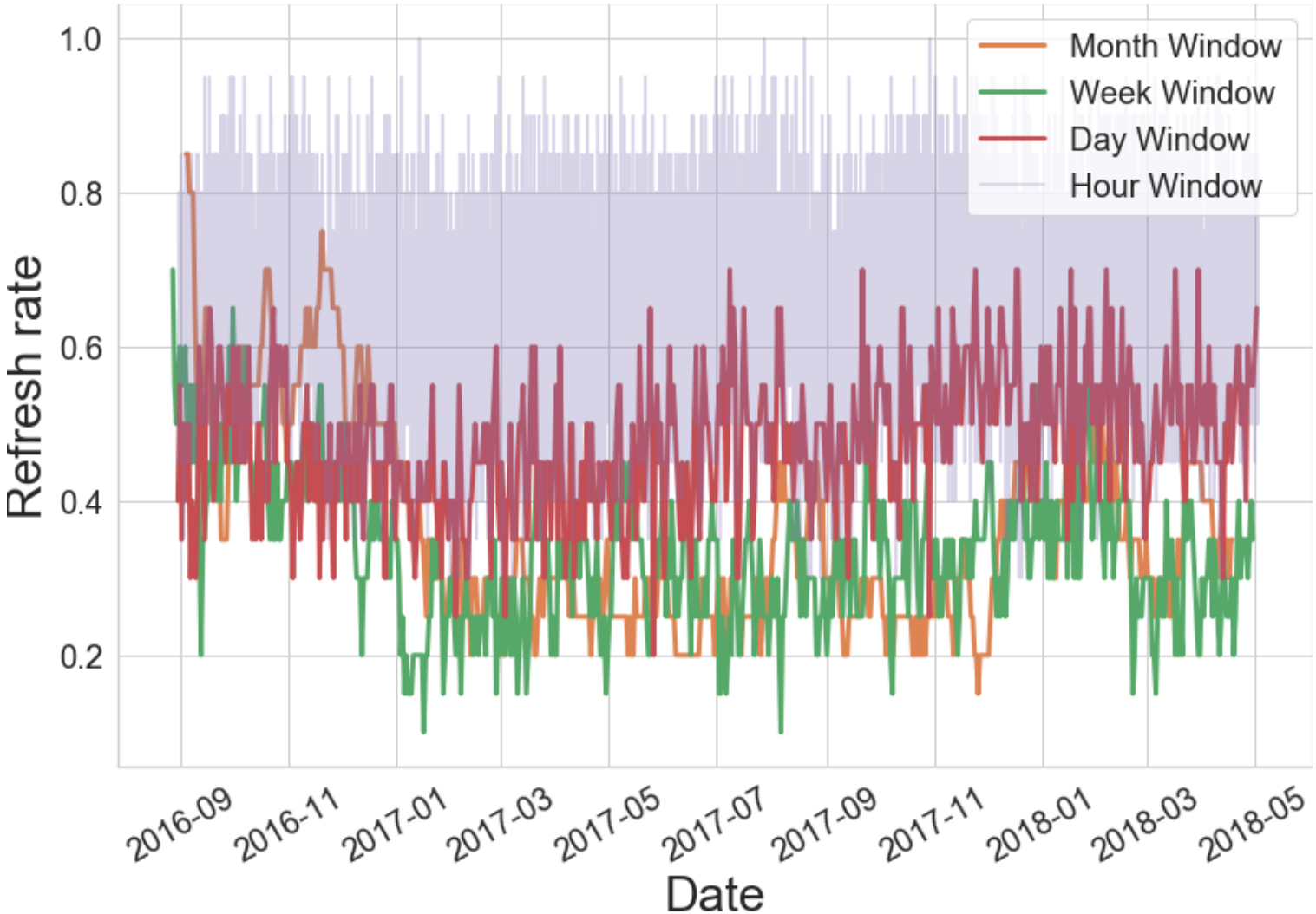
$$J(A, B) = \frac{A \cap B}{A \cup B}.$$

Refresh rate top N users, windows W_A, W_B

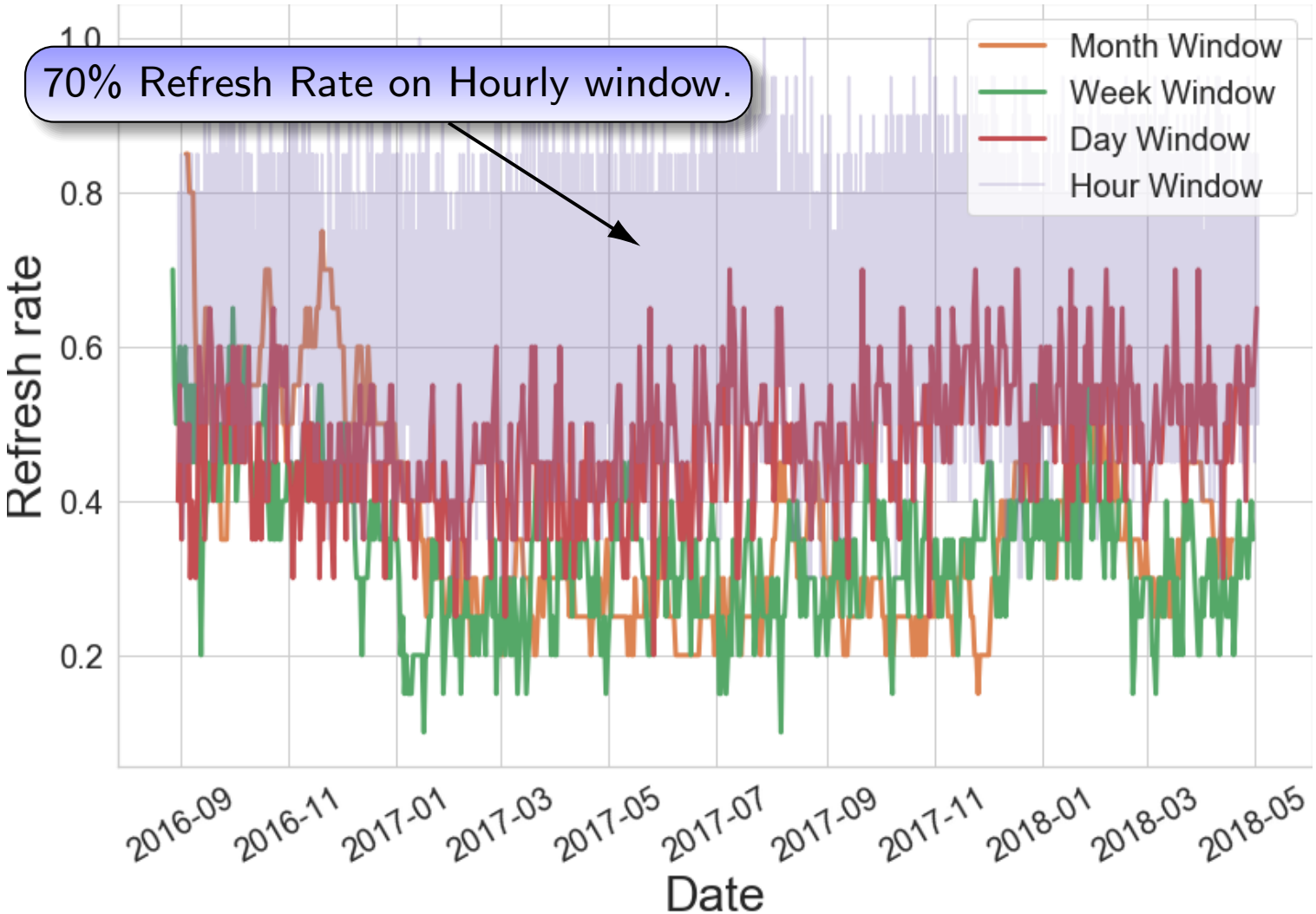
Refresh rate $R = 1 - J(A, B)$

$0 \rightarrow W_A$ same users as W_B and $1 \rightarrow$ no users the same.

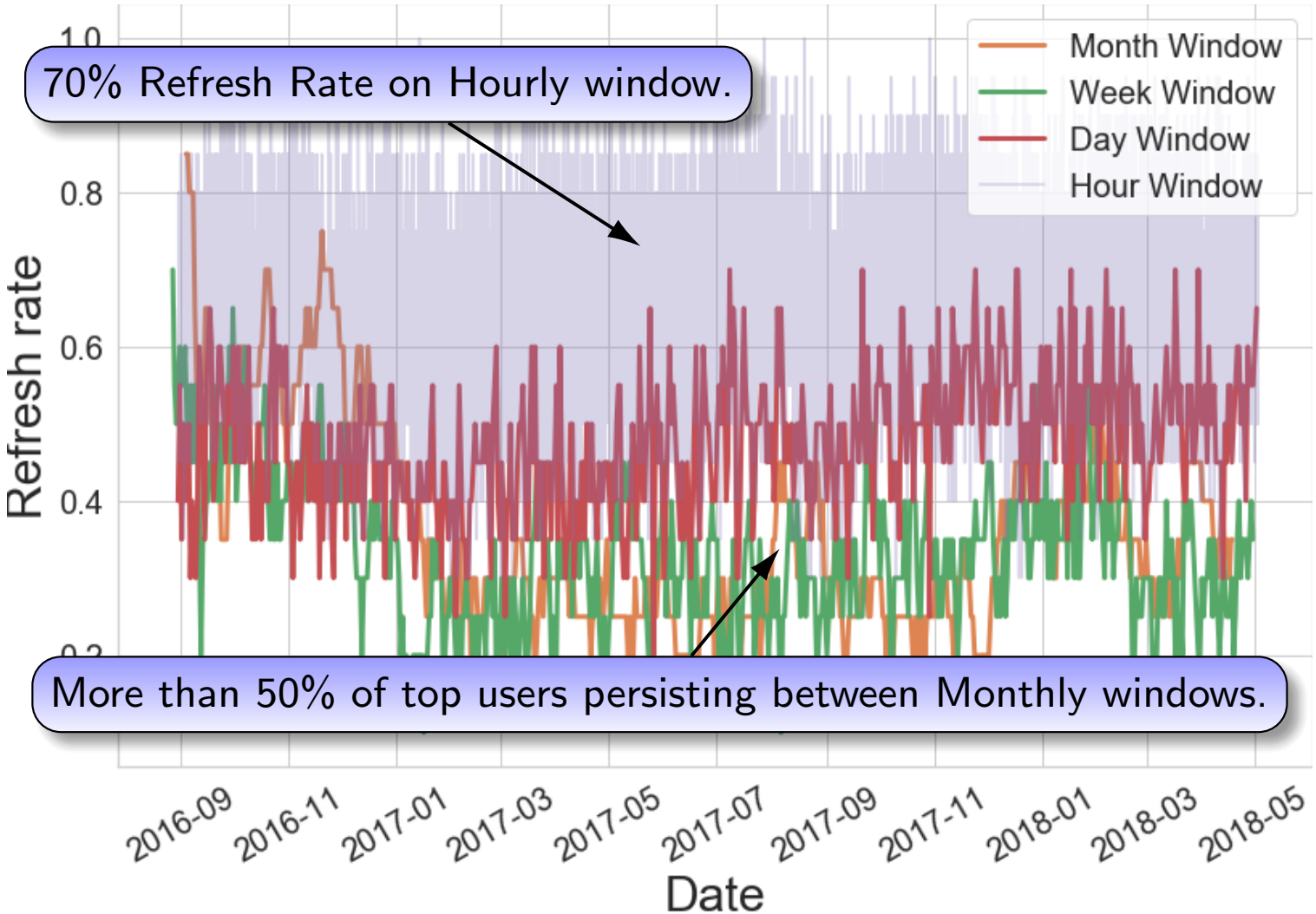
Is gab controlled by an "elite"?



Is gab controlled by an "elite"?



Is gab controlled by an "elite"?



Conclusions (about gab)

- Is gab growing? What drives this?
  Gab is evolving with rapid growths witnessed at small windows, driven by community specific events.

Conclusions (about gab)

- Is gab growing? What drives this?
  Gab is evolving with rapid growths witnessed at small windows, driven by community specific events.
- Gab is not a "social" network – interactions between "strangers" not friends.

Conclusions (about gab)

- Is gab growing? What drives this?
  Gab is evolving with rapid growths witnessed at small windows, driven by community specific events.
- Gab is not a "social" network – interactions between "strangers" not friends.
- The interactions within Gab do not always form a "connected" community. We observed a daily shattering of the LCC and diurnal change of regime never observed before.

Conclusions (about gab)

- Is gab growing? What drives this?
  Gab is evolving with rapid growths witnessed at small windows, driven by community specific events.
- Gab is not a "social" network – interactions between "strangers" not friends.
- The interactions within Gab do not always form a "connected" community. We observed a daily shattering of the LCC and diurnal change of regime never observed before.
- Are a cadre of elite users controlling users' attention?
  Not clear: At longer timescales there is a group who receives a lot of attention. At only one timescale you will get a misleading answer.

Conclusions (about Raphtory and temporal networks)

- Temporal networks provide a rich array of techniques that can get more out of your network data.
  - Window based analysis provides many levels of insight.

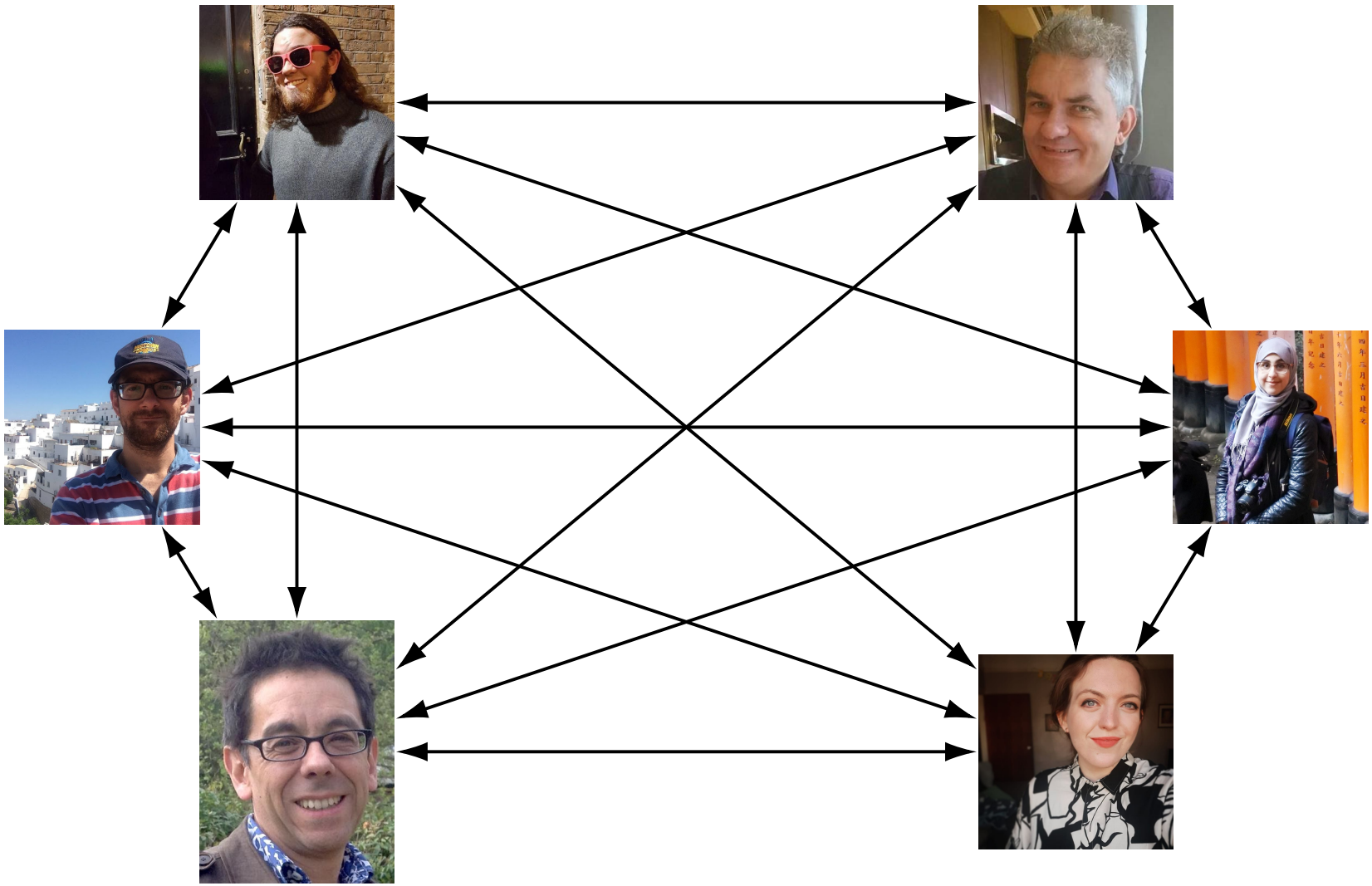
Conclusions (about Raphtory and temporal networks)

- Temporal networks provide a rich array of techniques that can get more out of your network data.
  - Window based analysis provides many levels of insight.
- The Raphtory tool is a great way to look at temporal graph data.

Conclusions (about Raphtory and temporal networks)

- Temporal networks provide a rich array of techniques that can get more out of your network data.
  - Window based analysis provides many levels of insight.
- The Raphtory tool is a great way to look at temporal graph data.
  - Urban analytics – intervention in social networks.
  - Bitcoin/blockchain – tracking "dark markets".
  - Semantic networks – changing word meanings.
  - Other social networks – compare and learn more.

Our Raphtory social network (a small subgraph)



Our Raphtory social network (a small subgraph)