# Social networks and scripts

Zvi Lotker

**IRIF-France** 

### Why study algorithm in literature

- In the Present we can't read all! So In the future machines will have to read for us!
- 2. Social network and science
- We would like the network to tale a story
   In literature we get the signal without the noise
   "All the world's a stage, And all the men and women merely players; As You Like It, Act II, Scene VII





#### Why study algorithm in literature

- It is an more fun to study literary than Facebook
- 2. I had a dream when I was a kids





#### How to study scripts

• By Social networks.

#### How Social network look



#### Do they

#### How literature networks look



Do they

#### How literature networks looks



#### Do they

#### What Questions are interesting in literature

- How are the stars of the story?
- How are the good and bad?





literature networks













#### From scripts to graph

#### • THREEPIO

- Did you hear that? They've shut down
- the main reactor. We'll be destroyed
- for sure. This is madness!
- Rebel troopers rush past the robots and take up positions in
- the main passageway. They aim their weapons toward the door.

#### THREEPIO

- We're doomed!
- LUKE
- Hurry up! Come with me! What are you
- waiting for?! Get in gear!





• The beginning of the movie 0-30

• The end of the movie -20,-1







#### Good vs. Bad

 Tell me who you friends are and tell you who you are



## Problems of validating algorithms in social networks

- 1. Privacy -> Anonymity
- 2. Avoiding open conflict
- 3. People's opinions are fluid



#### How to overcome these problems?

- 1. Build a machine to prove it!
  - Galileo Galilee
  - The Large Hadron Collider
  - Electroencephalography
- 2. Use the collective memory
  - Sigmund Freud







#### Julius Caesar for dummies

#### JULIUS CAESAR (in 3 Panels)

Cassius and Brutus assassinate Julius Caesar.



Marc Antony gives a really big speech.



Cassius and Brutus kill themselves.



www.goodticklebrain.com

©2014 Mya Gosling

#### Julius Caesar by William Shakespeare Act 4









#### Julius Caesar by William Shakespeare Act 5

#### • "This was the noblest Roman of them all"





## "Big Brother is Watching You"

- "Tell me who your friends are, and I
  - will tell you who you are"

#### • Or

• "Tell me who your friends vote for, and I will tell you to who you'll vote for"





The social network of Casca

## "Big Brother is Watching You"

- "Tell me who your friends are, and I
  - will tell you who you are"

#### • Or

 "Tell me who your friends vote for, and I will tell you to who you'll vote for"



#### Collapse operation

#### Collapsed a collection of vertices v1,v2,...vk in to a single vertex



#### Collapse operation

• Collapsed a collection of vertices Cassius and Brutus in to a single vertex Pindarus Messenger Tintinius Young Cato Antony Strato Octavius Dardanius Brutus First Soldier First Soldier Lucilius Strato Second Se Volumnius Second Soldier Clitus Young Cato Dardanius Pindarus Messenger Tintinius Young Cato Antony Octavius rutur First Soldier Lucilius Strate Volumnius Second Soldier Clitus Dardanius



#### Absorbing operation

• removing all outgoing edges





#### Absorbing operation

#### removing all outgoing edges



#### Random walks



- Random step:
  - Move to an adjacent node chosen at random (and uniformly)
- Random walk:
  - Take an infinite sequence of random steps



#### The voting algorithm

1. Collapse the ANCHORS R to a single node and

Collapsed the ANCHORS B to a single node.

- 2. Make both Node R and B Absorbing.
- 3. Compute the hitting probability to hit R.



#### How do we find the anchors?

 In the case of *Julius Caesar* by William Shakespeare we look on the four act.

2. In the general case and more complicated





How do we compute the hitting time?

1. Node v which belongs to the red anchors set,

- $R_v = 1$
- 2. Node v which belongs to the Blue anchors set,
  - $R_v = 0$

3. The equation of the unknown states

• 
$$R_{v} = \frac{1}{|N_{n}|} \sum_{u \in N_{n}} R_{u}$$

#### The results of the voting algorithm in *Julisus Caesar* The Complete Play



#### The results of the voting algorithm in *Julisus Caesar* The 5 act



#### Election LUKE v.s VADER





#### Election LUKE v.s VADER







• How does time move in a social network?

• Are there several time dimensions in a social network?

 How can we detect a crucial event in the evolution of a social network?

#### Heraclitus

- Panta rhei, "everything flows"
- A Social network changes in time

• 'nature loves to hide' Heraclitus



## A principle of measure time.



- A useful guideline to understanding how time flows is to ask: what changes?
- When trying to capture the meaning of time in social networks, the first observation is that in social networks, time is discrete
- The change of the structure in social networks is either adding or deleting a vertex/edge, which are atomic events.





- We consider two networks, both have 10 nodes, in each round a single edge a is arrive.
- In one network have a single community
- In the second we have two community



• Is it possible to distinguish between the two cases



#### An example that emphasizes the problem

• Consider two different cases





#### An example that emphasizes the problem

• Consider two different cases





#### An example that emphasizes the problem

#### • Consider two different cases



#### Time in theatre



- The time of the text: from act I to the end
- The time of the show: about 3-5 hours.
- The historic time, or the time the play takes place in (the 16th century in Hamlet, more or less).
- The time of the story.

 Time perception: is a field of study within <u>psychology</u> and <u>neuroscience</u> that refers to the subjective experience of <u>time</u>.

• For **dogs** time flows faster than for humans

• For **Babies** time flows slower





0.20.05

1-90 lb (24-40 kg

Dog years



- Time perception: is a field of study within <u>psychology</u> and <u>neuroscience</u> that refers to the subjective experience of <u>time</u>.
- When we get **older** time flows faster.









- Time perception: is a field of study within <u>psychology</u> and <u>neuroscience</u> that refers to the subjective experience of <u>time</u>.
- The oddball effect





- How does it work?
  - We have a clock in our mind.
  - This is not an accurate clock.
  - This is an important clock.
- So when our clock drifts it is an indication of an important event.

• So we can use Time perception to find important event.

#### From time to clock

- Usually if we have two clocks we like to synchronize the clocks
- But In our case we like to find the point that the clocks are not synchronized.



8:00:00 8:00:00

JAN 01, 2016 8:00:00



8:00:00 8:00:24

JAN 21 2016 8:00:00

#### What does it mean to synchronize clocks



#### What are clocks?



• Def: A **clock** is an instrument to indicate, keep, and co-ordinate <u>time</u>

• We are interested in the imprecise clocks

- Def: A Clock is a strictly monotonic Increasing continuous function.
  - $C: [a, b] \to \mathbb{R}$

## Property of clock

- Def: The standard clock is the identity function
- For each clock C there exists its own inverse function  $C^{-1}$ , which is also a clock.
- A discrete clock is a discrete monotonic increasing function over the real numbers  $C: A \to \mathbb{R}$ , where the set  $A \subset \mathbb{R}$  is discrete.
- Lemma: every discrete clock can be extended to a clock.

#### When two clocks are linear

- Def: The two clocks C<sub>1</sub>,C<sub>2</sub> are linear with respect to each other, if there exist two linear transformations
  - $L_1, L_2$  s.t  $C_1(L_1(t)) = L_2(C_2(t))$  as a function.
- This is an equivalent relation.
- Why we need two linear functions and not one?
  - One is for normalizing the x coordinate and one to normalize the y coordinate

#### All clocks are normalize

- Lemma Let C<sub>1</sub> be a clock that is defined on the time interval [a, b]. There exists a unique normalized clock C<sub>n</sub>, such that C<sub>n</sub> is equivalent to C<sub>1</sub>
- Proof:
- We first normalize the time [a,b] to do so we define the clock
  - $C_2(t) = C_1(a + (a b)t)$  note that is a linear function



#### The Second Moment Clock diagram

- Since we normalize the clocks we can look at the function
- $S(c_1,c_2)(t)=c_2(t) c_1(t)$
- Note that
  - $S(c_1, c_2)(0) = S(c_1, c_2)(1) = 0$



#### Clock Drift

Def: of clock drift Two clocks, C<sub>1</sub>,C<sub>2</sub>, defined on the set A, have a clocks drift with respect to each other, if they are not equivalent on the set A.



#### Single Clock Drift

• Def: Single clock drift The clocks  $C_1, C_2$  have a single clock drift on set A if there exists  $c \in A$ , such that the clocks  $C_1; C_2$  are equivalent on sets  $A_1 = \{x \in A | x \le c\}$  and on the set  $A_2 = \{x \in A | x > c\}$ , but they are not equivalent on A. In this case, c is the beginning of the clock drift.



#### How to model social networks with time?

- We want to use time So we look on Evolving graphs
- We are giving an link Stream,
  - Link Stream = {(( $v_i$ , ui), ti)  $|v, u \in V, ti \in \mathbb{R}$ }
  - Weighted Link Stream = {( $(v_i, ui), t_i, w_i$ ) |  $v, u \in V, ti \in \mathbb{R}, wi \in \mathbb{R}$ }
- A useful analog is to consider static social network as image and evolving social network as movie

#### From Weighted Link Stream to two clocks

- The event clock C<sub>e</sub> is the number of edges arriving up to time t in the evolving graph.
- The weighted clock C<sub>w</sub> is the weighted of all the edges arriving up to time t.

#### The Correlation Between Two Clocks Example

• Consider The Julius Caesar  $\rho(C_e, Cw) = 0.991631$ 



#### The gap method

- There are two ways to find the beginning of a drift in the case of a single drift
- The Gap method.
- $g(c_1, c_2) = \max_{0 \le t_1 \le 1} c_1(t_1) c_2(t_1) \min_{0 \le t_2 \le 1} c_1(t_2) c_2(t_2)$
- Note that  $g(c_1, c_2) = g(c_2, c_1)$
- We denote the

• 
$$T_M = Arg \max_{0 \le t_1 \le 1} c_1(t_1) - c_2(t_1),$$
  
•  $T_m = Arg \min_{0 \le t_1 \le 1} c_1(t_1) - c_2(t_1)$ 

#### The dynamic programing method

- We can use dynamic programing to find the beginning of the deft.
  - $\min_{a_1,a_2,m} \sum_{i=1}^{m} \left( C_1(t_i) a_1 C_2(t_i) \right)^2 + \sum_{i=m+1}^{n} \left( C_1(t_i) \left( a_2 C_2(t_i) + 1 a_2 \right) \right)^2$
  - S.t
  - *a*<sub>1</sub>, *a*<sub>2</sub>>0
  - $a_1 C_2(t_m) \le a_2 C_2(t_{m+1}) + 1 a_2$

#### Comparing the methods

- Theorem: Let C<sub>1</sub>,C<sub>2</sub> be two continuous clocks, with a single clock drift on the time interval [0; 1]. Assume that T is the beginning of the drift, then one of the alternatives holds:
  - If for all 0 < t < T,  $C_1(t) < C_2(t)$ , then  $T_M = T_{DP} = T$ .
  - If for all 0 < t < T,  $C_1(t) > C_2(t)$ , then  $T_m = T_{DP} = T$ .
- Claim: when we have two clocks C<sub>e</sub>, C<sub>w</sub> with a single clock drift the beginning of a drift is an important event in the social network

#### Othello



#### Romeo Juliet

ROMEO



#### Julius Caesar



#### Thanks





