Evolution over time of the structure of social graphs

Student Leonardo Serilli
Supervisors Frédéric Giroire, Nicolas Nisse, Malgorzata Sulkowska
Year 2021/2022

TABLE OF CONTENTS

01

INTRODUCTION

Context
Scope of the project
Data

CONTEXT

- **Scopus** Database

CONTEXT

- **Scopus** Database

CONTEXT

- **Scopus** Database
- **250K** computer science **authors**
- **Collaborations** (**co-authorship** relationship) between **1990 and 2018**

CONTEXT

- **Scopus** Database
- **250K** computer science **authors**
- **Collaborations** (**co-authorship** relationship) between **1990 and 2018**

Scopus ID

Total # collaborations (until 2014)

publications years

	ID	1990	1991	1992	1993	1994	1995	1996	1997	1998	...	2014	2015	2016	2017	2018	start_year	end_year
118063	26421678500	0	0	0	0	0	0	0	0	0	...	4	4	4	4	4	2014	2014
180546	56230251900	0	0	0	0	0	0	0	0	0	...	3	3	3	3	3	2013	2014
68772	7801413223	0	0	0	0	0	0	0	0	0	...	3	3	3	3	3	2011	2011
25152	6603158006	0	0	0	0	0	0	0	0	0	...	0	0	0	0	4	2018	2018
96494	20434297300	0	0	0	0	0	0	0	0	0	...	16	16	16	16	16	2013	2013
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
228654	57203927130	0	0	0	1	1	1	1	1	1	...	31	32	39	47	48	1991	2018
115362	25647427000	0	0	0	0	0	0	0	0	0	...	0	0	0	0	49	2018	2018
176446	56066133100	0	0	0	0	0	0	0	0	0	...	0	0	0	3	3	2017	2017
64352	7202888402	0	0	0	0	0	0	0	0	0	...	3	3	3	3	3	2006	2006
101801	23099287300	0	0	0	0	0	0	0	0	0	...	12	17	17	17	22	2009	2018

232838 rows × 35 columns

SCOPE OF THE PROJECT

- Build the **collaboration network G=(V,E)** **| V = {authors}, E = {collaborations}.**

SCOPE OF THE PROJECT

- Build the **collaboration network G=(V,E) | V = {authors}, E = {collaborations}.**
- Study some of it's mathematical **properties** and their **evolution over time.**

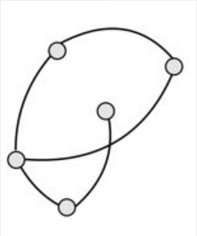
SCOPE OF THE PROJECT

- Build the **collaboration network G=(V,E) | V = {authors}, E = {collaborations}.**
- Study some of it's mathematical **properties** and their **evolution over time**.
- Implement a **model** to **generate** similar graphs and so able to **predict** their evolution .

SCOPE OF THE PROJECT

- Build the **collaboration network G=(V,E) | V = {authors}, E = {collaborations}.**
- Study some of it's mathematical **properties** and their **evolution over time**.
- Implement a **model** to **generate** similar graphs and so able to **predict** their evolution .

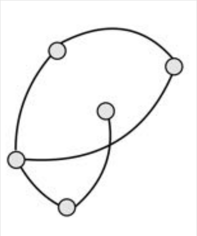
G_{1990}



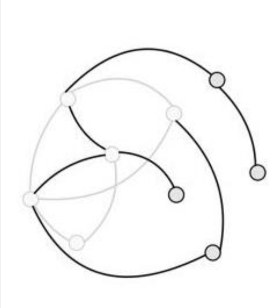
SCOPE OF THE PROJECT

- Build the **collaboration network G=(V,E) | V = {authors}, E = {collaborations}.**
- Study some of it's mathematical **properties** and their **evolution over time**.
- Implement a **model** to **generate** similar graphs and so able to **predict** their evolution .

G_{1990}

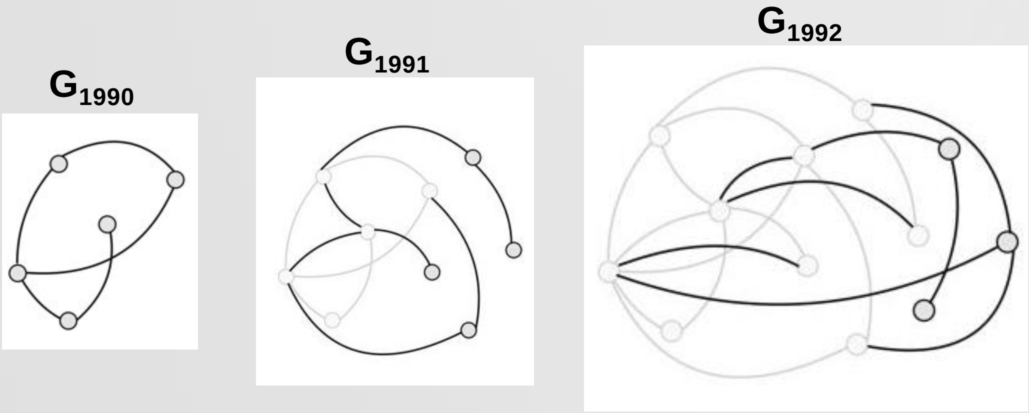


G_{1991}



SCOPE OF THE PROJECT

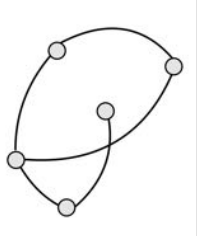
- Build the **collaboration network G=(V,E) | V = {authors}, E = {collaborations}.**
- Study some of it's mathematical **properties** and their **evolution over time**.
- Implement a **model** to **generate** similar graphs and so able to **predict** their evolution .



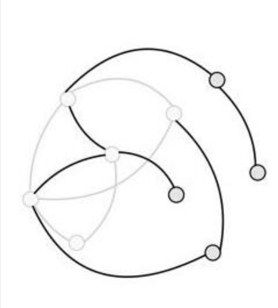
SCOPE OF THE PROJECT

- Build the **collaboration network G=(V,E) | V = {authors}, E = {collaborations}.**
- Study some of it's mathematical **properties** and their **evolution over time**.
- Implement a **model** to **generate** similar graphs and so able to **predict** their evolution .

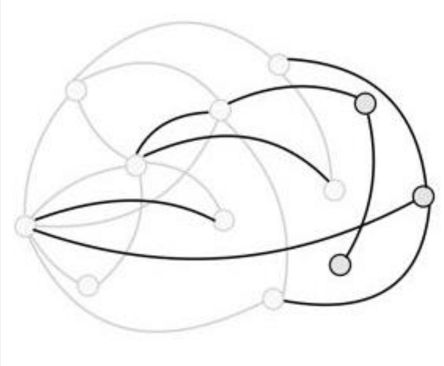
G_{1990}



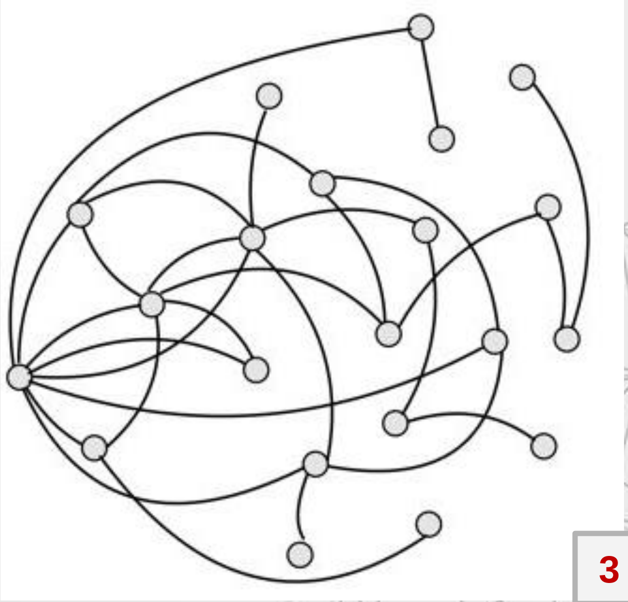
G_{1991}



G_{1992}



G_{2018}



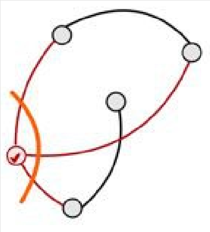
RELEVANT PROPERTIES

- **Vertex trajectory**: **$d_v(y)$** degree of **v** at the given year **y.**

RELEVANT PROPERTIES

- **Vertex trajectory**: $d_v(y)$ degree of **v** at the given year **y.**

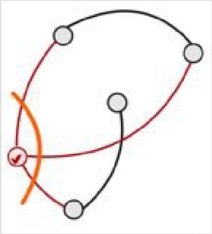
dv(1990)= 3



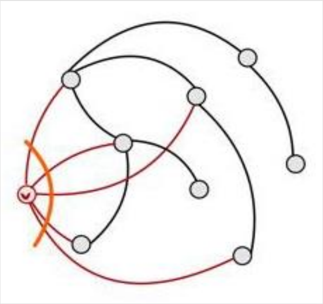
RELEVANT PROPERTIES

- **Vertex trajectory**: $d_v(y)$ degree of **v** at the given year **y.**

dv(1990)= 3



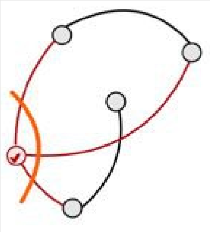
dv(1991)= 5



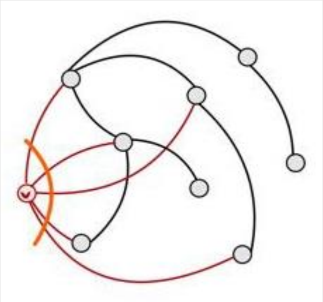
RELEVANT PROPERTIES

- **Vertex trajectory**: $d_v(y)$ degree of **v** at the given year **y.**

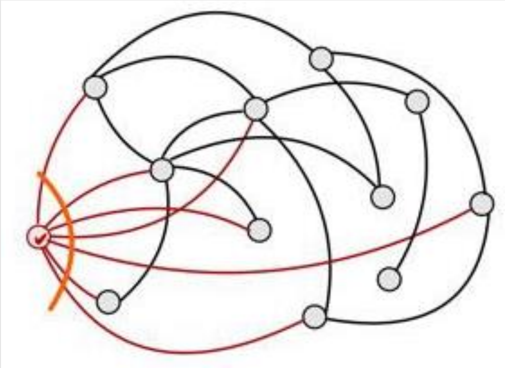
dv(1990)= 3



dv(1991)= 5



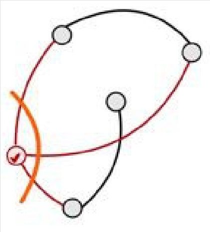
dv(1992)= 7



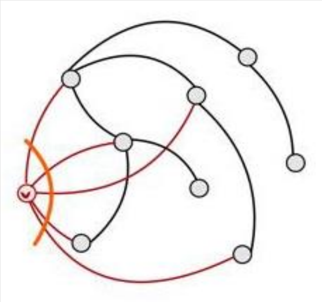
RELEVANT PROPERTIES

- **Vertex trajectory**: $d_v(y)$ degree of **v** at the given year **y.**

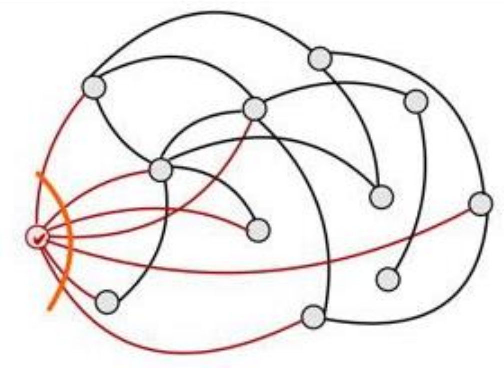
dv(1990)= 3



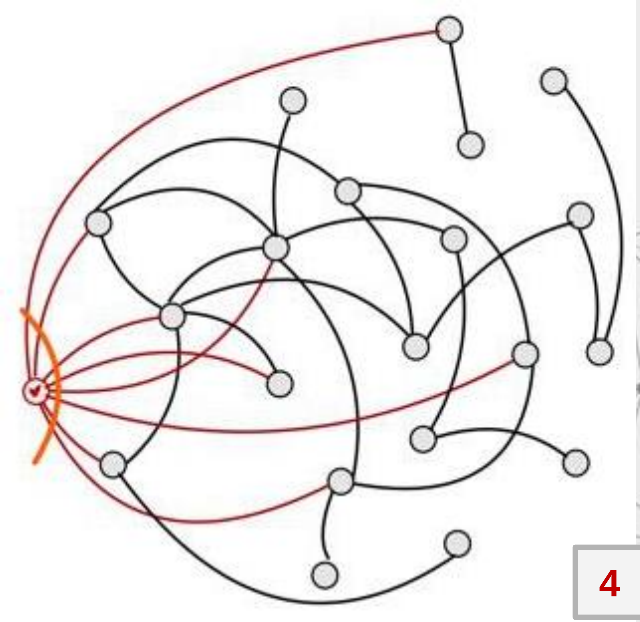
dv(1991)= 5



dv(1992)= 7

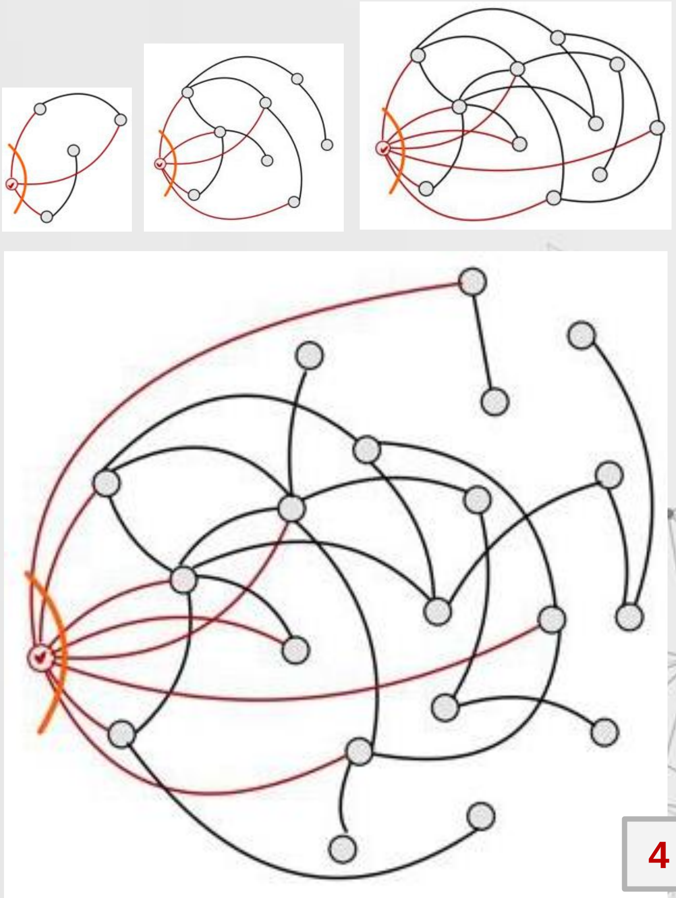


dv(2018)= 8



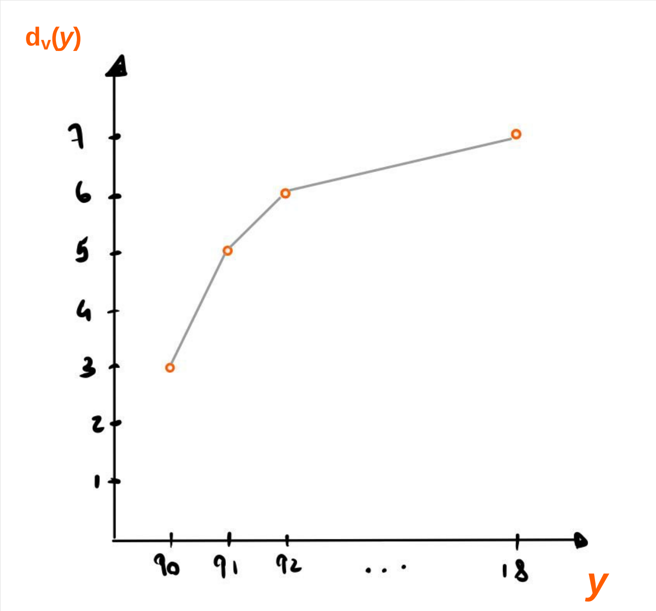
RELEVANT PROPERTIES

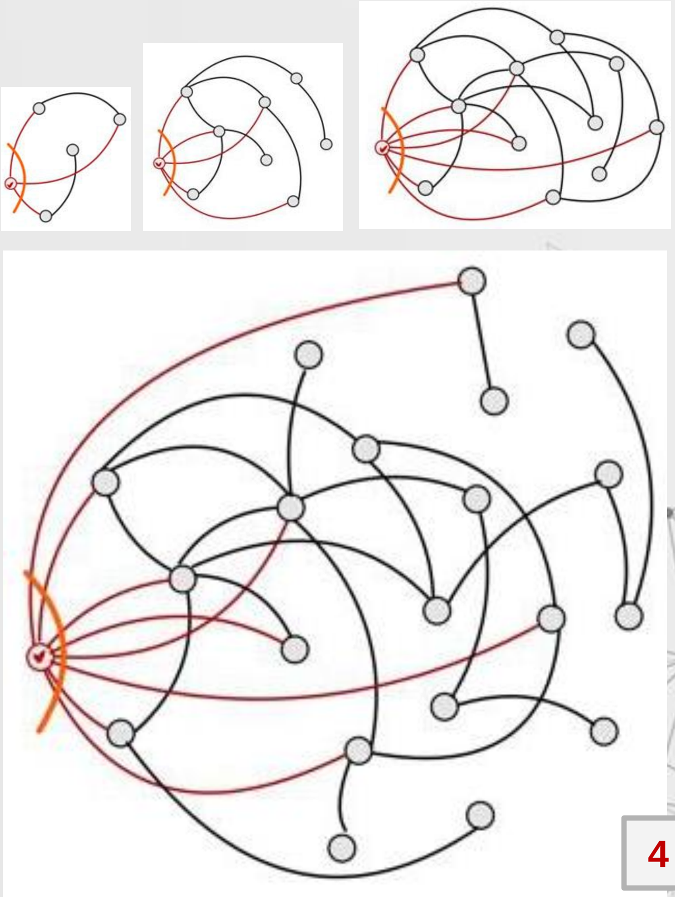
- **Vertex trajectory**: $d_v(y)$ degree of **v** at the given year **y.**



RELEVANT PROPERTIES

- **Vertex trajectory**: $d_v(y)$ degree of **v** at the given year **y.**





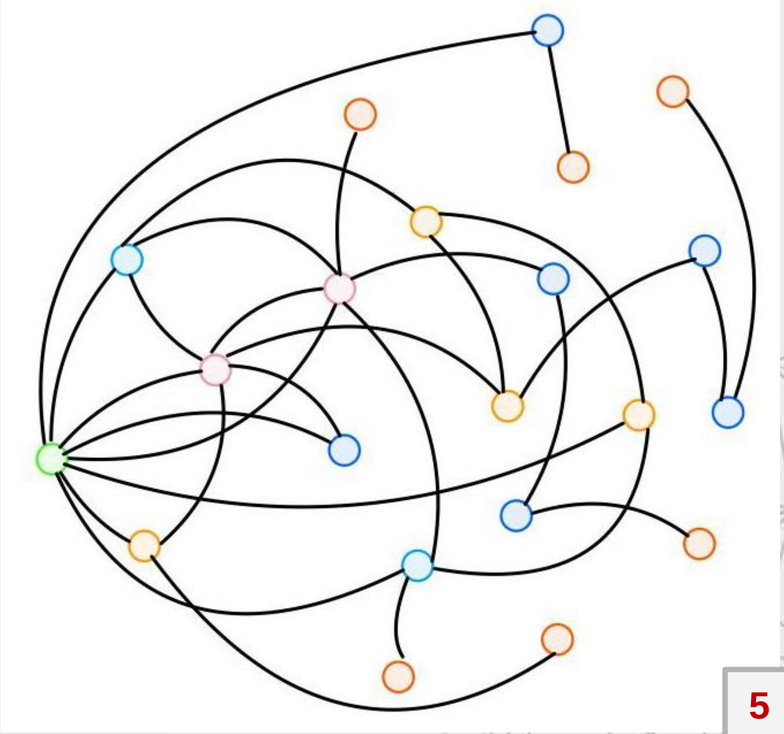
RELEVANT PROPERTIES

- **Vertex trajectory**: $d_v(y)$ degree of **v** at the given year **y.**
- **Degree Distribution**: $\mathbf{M_k = n_k/n}$ **|** $\mathbf{n_k}$ **=** #nodes of **degree k** in G.

RELEVANT PROPERTIES

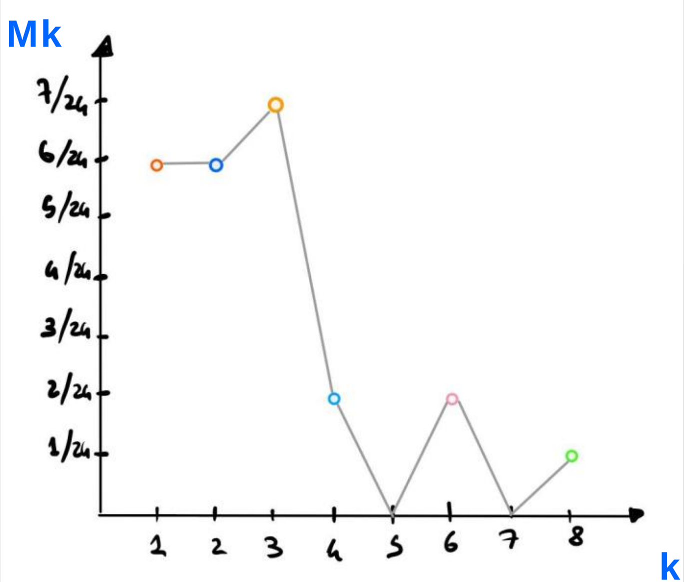
- **Vertex trajectory**: $d_v(y)$ degree of **v** at the given year **y.**
- **Degree Distribution**: $M_k = n_k/n$ | n_k = #nodes of **degree k** in G.

G_{2018}

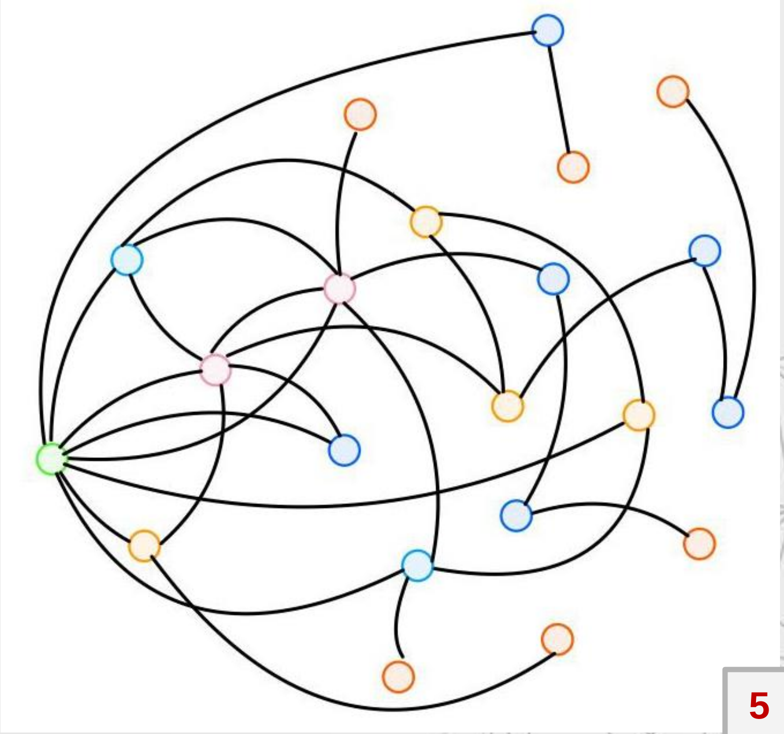


RELEVANT PROPERTIES

- **Vertex trajectory**: $d_v(y)$ degree of **v** at the given year **y.**
- **Degree Distribution**: $M_k = n_k/n$ **|** n_k **=** #nodes of **degree k** in G.



G_{2018}



02

STATE OF THE ART

Network evolution over time
Power law and scale free networks
Power law in real world networks
Theoretical model
Matching model

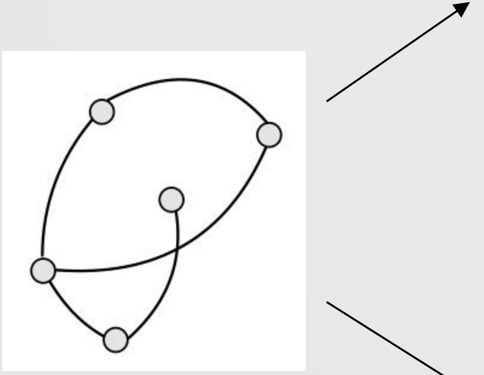
NETWORK EVOLUTION OVER TIME

- **G = (V, E)** evolves over time.

NETWORK EVOLUTION OVER TIME

- **G = (V, E)** evolves over time.

Two kinds of event:

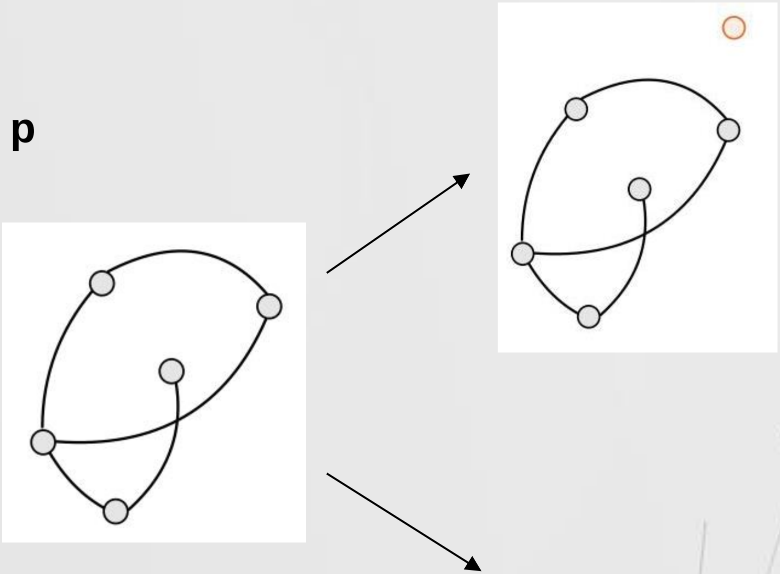


NETWORK EVOLUTION OVER TIME

- **G = (V, E)** evolves over time.

Two kinds of event:

- **Node + Edge event**, probability **p**

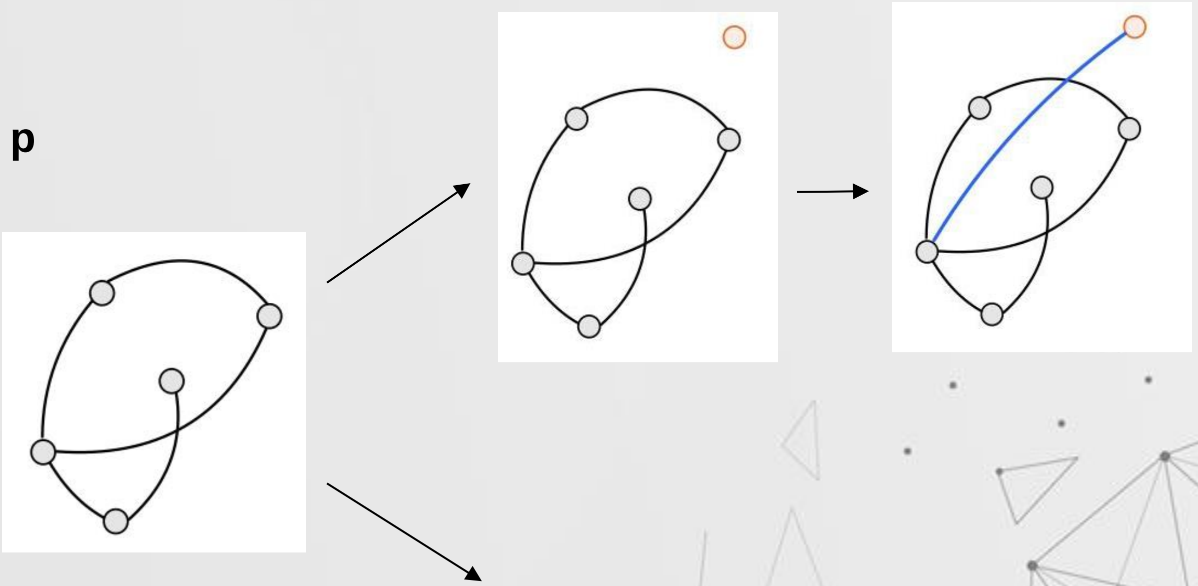


NETWORK EVOLUTION OVER TIME

- **G = (V, E)** evolves over time.

Two kinds of event:

- **Node + Edge event**, probability **p**

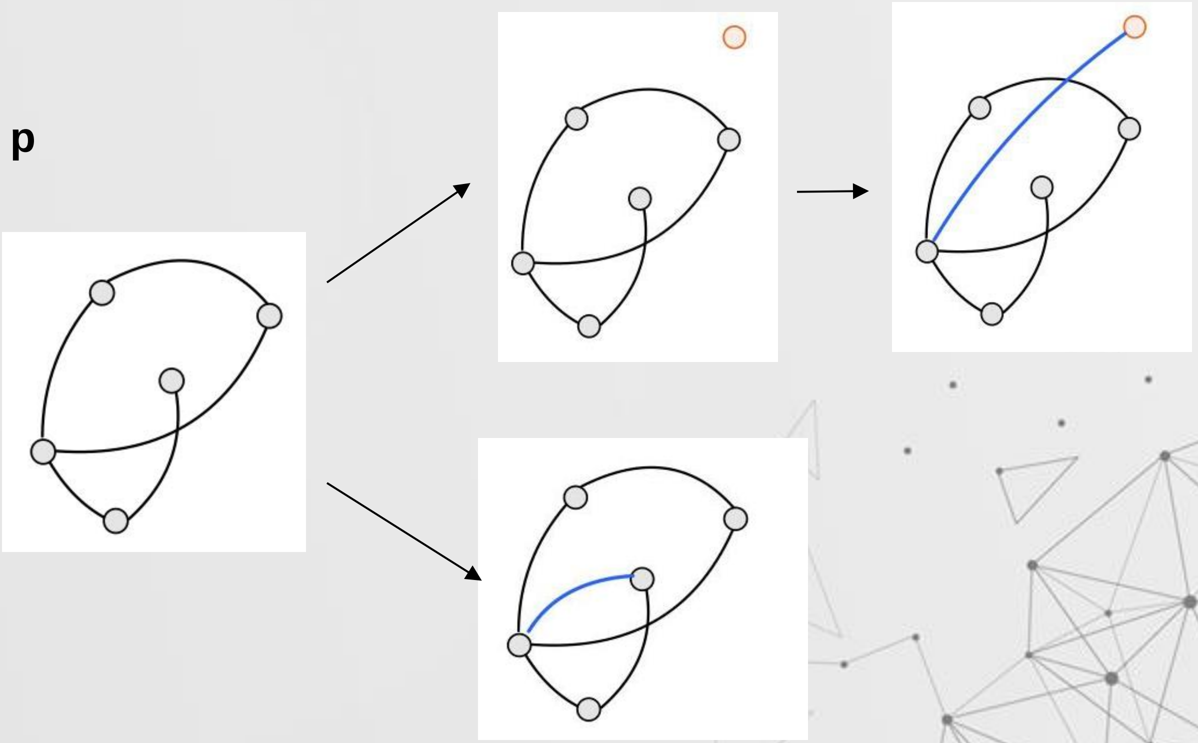


NETWORK EVOLUTION OVER TIME

- **G = (V, E)** evolves over time.

Two kinds of event:

- **Node + Edge event**, probability **p**
- **Edge event**, probability **(1-p)**

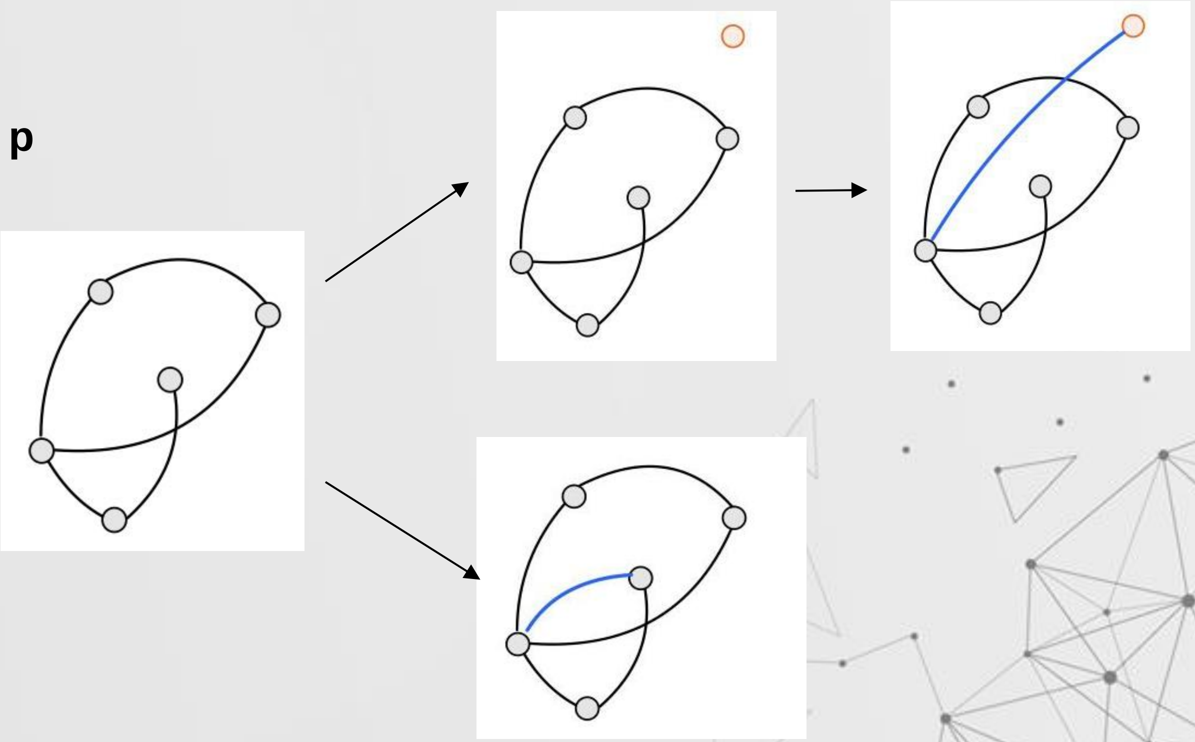


NETWORK EVOLUTION OVER TIME

- **G = (V, E)** evolves over time.

Two kinds of event:

- **Node + Edge event**, probability **p**
- **Edge event**, probability **(1-p)**



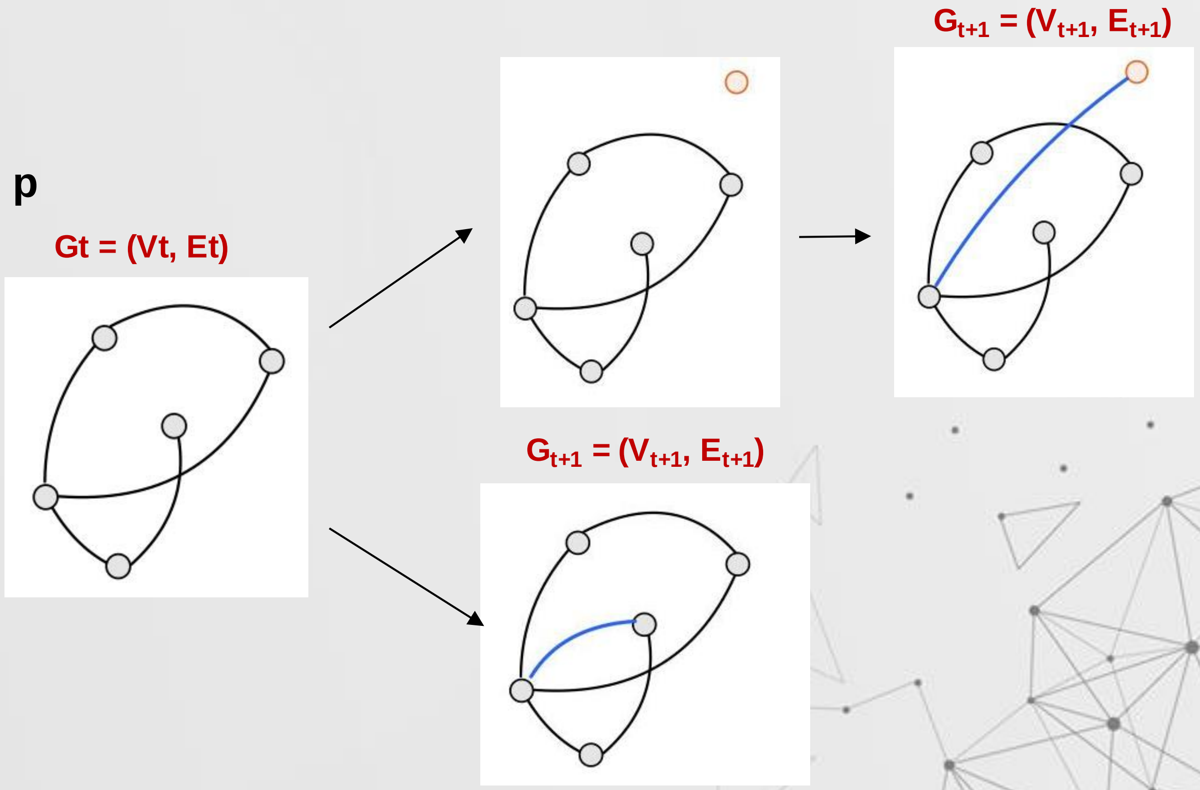
- Let **t∈N** - #events occurred
- $G_t = (V_t, E_t)$ graph after event **t**

NETWORK EVOLUTION OVER TIME

- **G = (V, E)** evolves over time.

Two kinds of event:

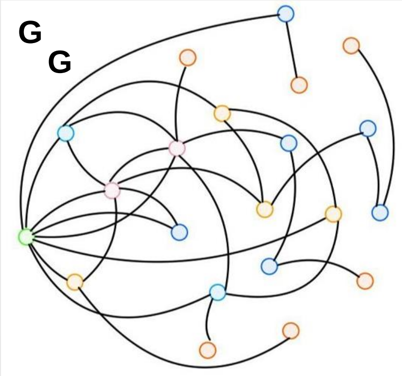
- **Node + Edge event**, probability **p**
- **Edge event**, probability **(1-p)**



- Let **t∈N** - #events occurred
- $G_t = (V_t, E_t)$ graph after event **t**

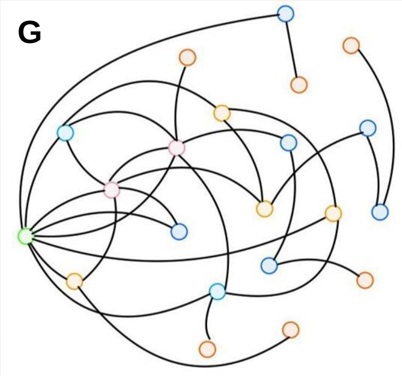
POWER LAW AND SCALE-FREE NETWORKS

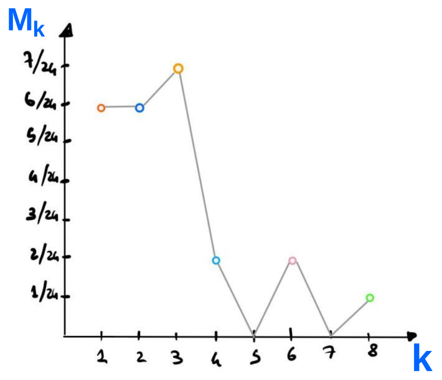
- **G = (V, E)** of **n** nodes



POWER LAW AND SCALE-FREE NETWORKS

- **G = (V, E)** of **n** nodes



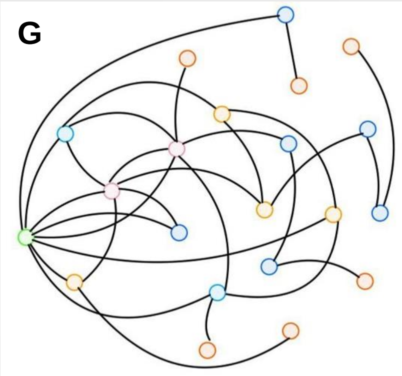


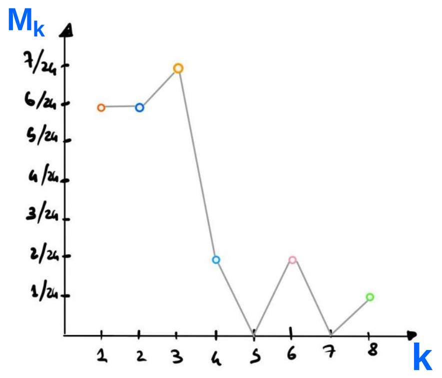
POWER LAW AND SCALE-FREE NETWORKS

- **G = (V, E)** of **n** nodes

.

- Degree Distribution $\mathbf{M_k = n_k / n}$ | $\mathbf{n_k}$ = #nodes of **degree k** in G.



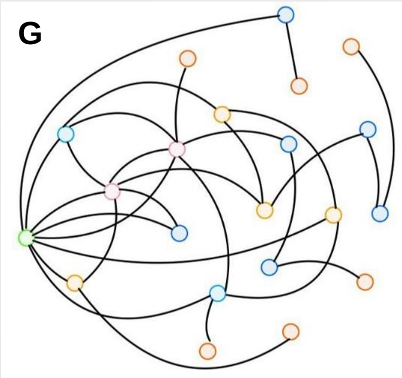


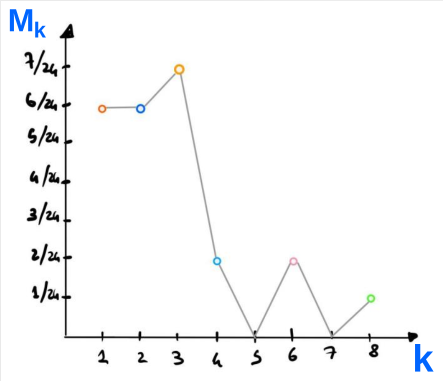
POWERLAW AND SCALE-FREE NETWORKS

- **G = (V, E)** of **n** nodes

- Degree Distribution $M_k = n_k / n$ | n_k = #nodes of **degree k** in G.

- A **scale-free** network respects the **power law** distribution $M_k \sim Ck^{-\lambda}$ | λ, $C > 0$.





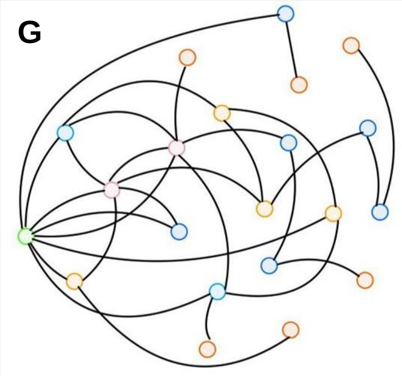
POWERLAW AND SCALE-FREE NETWORKS

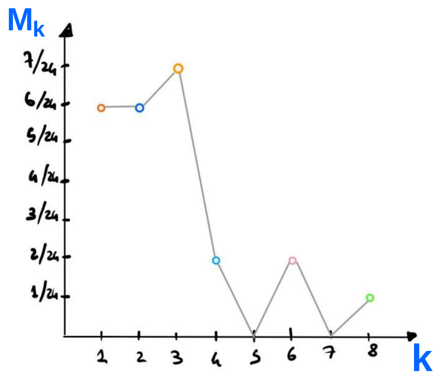
- **G = (V, E)** of **n** nodes

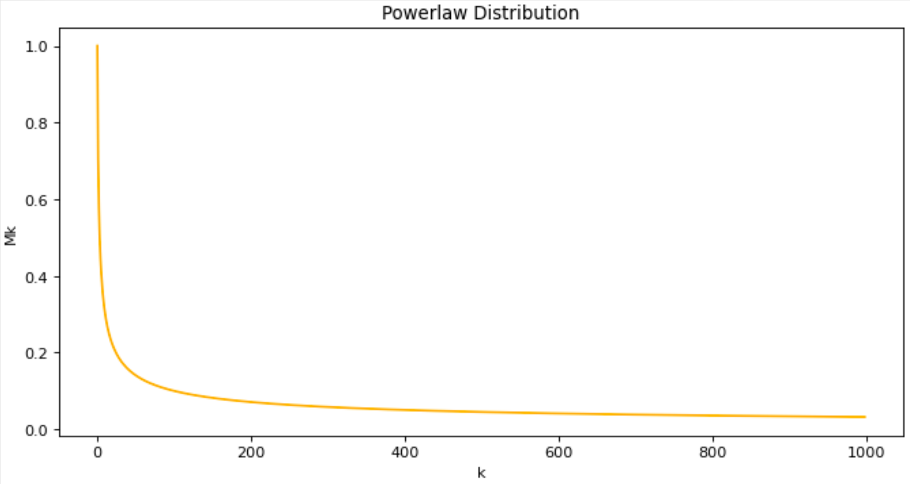
.

- Degree Distribution $\mathbf{M_k = n_k / n}$ | $\mathbf{n_k}$ = #nodes of **degree k** in G.

- A **scale-free** network respects the **power law** distribution $\mathbf{M_k \sim Ck^{-\lambda}}$ | $\boldsymbol{\lambda}$**, C** > 0.







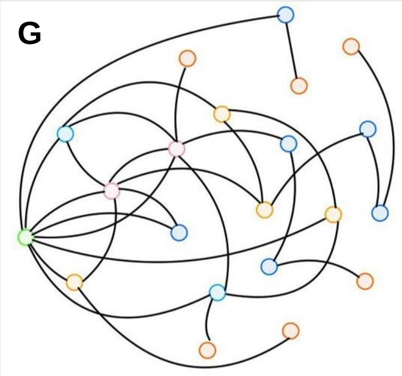
POWERLAW AND SCALE-FREE NETWORKS

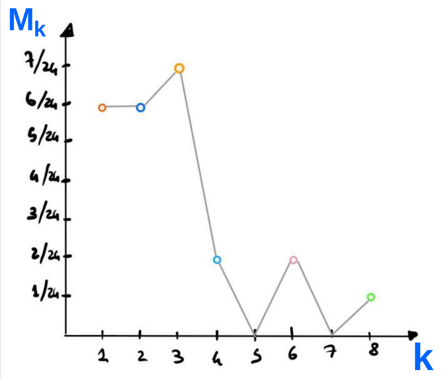
- **G = (V, E)** of **n** nodes

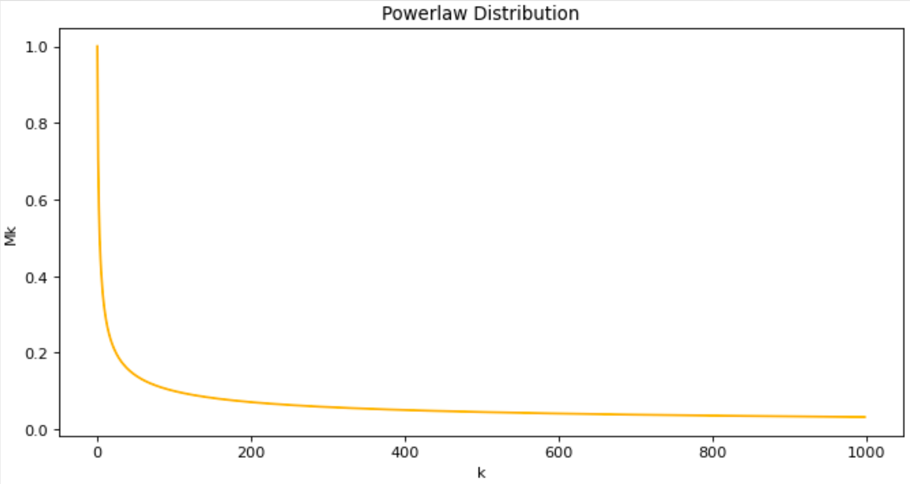
.

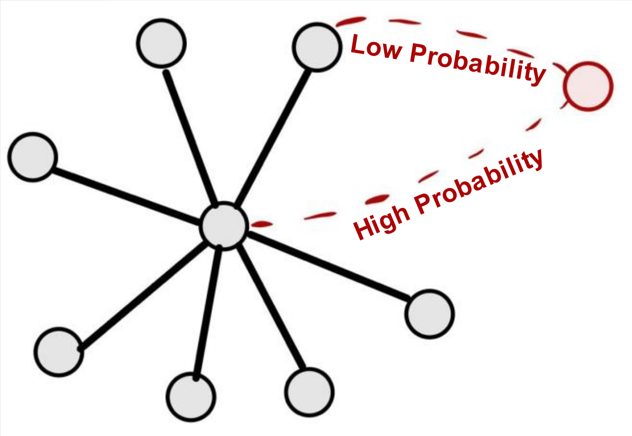
- Degree Distribution $\mathbf{M_k = n_k / n}$ **|** $\mathbf{n_k}$ **=** #nodes of **degree k** in G.

- A **scale-free** network respects the **power law** distribution $\mathbf{M_k \sim Ck^{-\lambda}}$ **| λ, C** > 0.









POWERLAW IN REAL WORLD NETWORKS

- Is often believed that most real-world networks are scale-free [1], [2], [3].

POWERLAW IN REAL WORLD NETWORKS

- Is often believed that most real-world networks are scale-free [1], [2], [3].

References:
- *[1] Derek J. de Solla P.* ***Networks of scientific papers****, 1965, doi:10.1126/science.149.3683.510.*
- *[2] Michalis F., Petros F., and Christos F.* ***On power-law relationships of the Internet topolog****y, 1999, doi:10.1145/316188.316229.*
- *[3] Bollobás B. and Riordan O****. Handbook of Graphs and Networks: From the Genome to the Internet,*** *2003. Pages 1–34.*

POWERLAW IN REAL WORLD NETWORKS

- Is often believed that most real-world networks are scale-free [1], [2], [3].

Powerlaw $M_k \sim Ck^{-\lambda}$

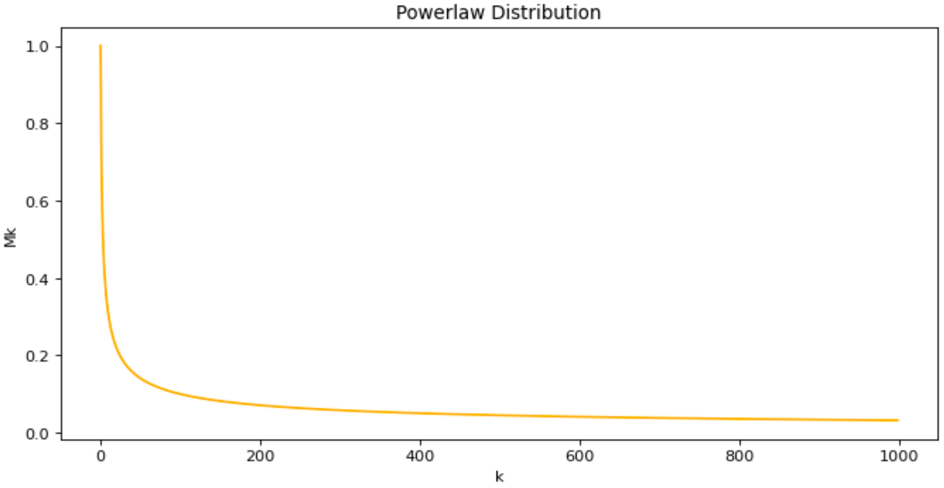
References:

- *[1] Derek J. de Solla P. **Networks of scientific papers**, 1965, doi:10.1126/science.149.3683.510.*
- *[2] Michalis F., Petros F., and Christos F. **On power-law relationships of the Internet topolog**y, 1999, doi:10.1145/316188.316229.*
- *[3] Bollobás B. and Riordan O**. Handbook of Graphs and Networks: From the Genome to the Internet,** 2003. Pages 1–34.*

POWERLAW IN REAL WORLD NETWORKS

- Is often believed that most real-world networks are scale-free [1], [2], [3].

Powerlaw $M_k \sim Ck^{-\lambda}$



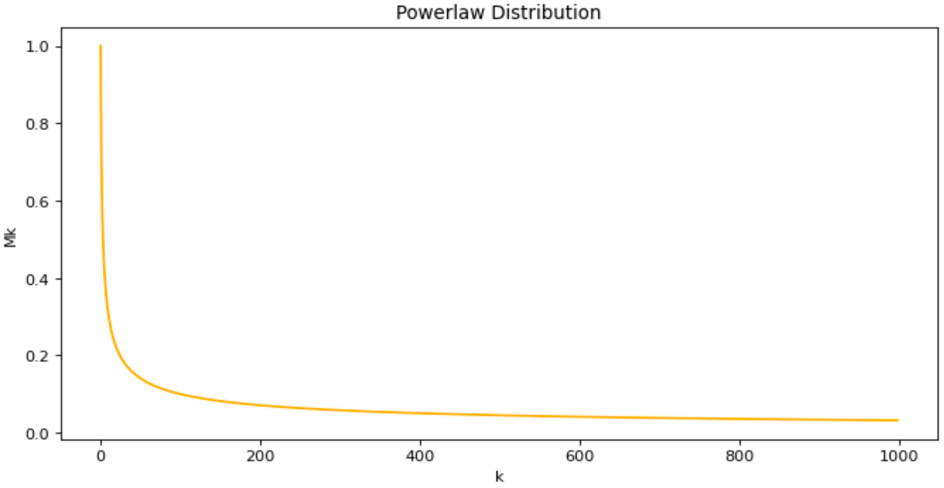
References:

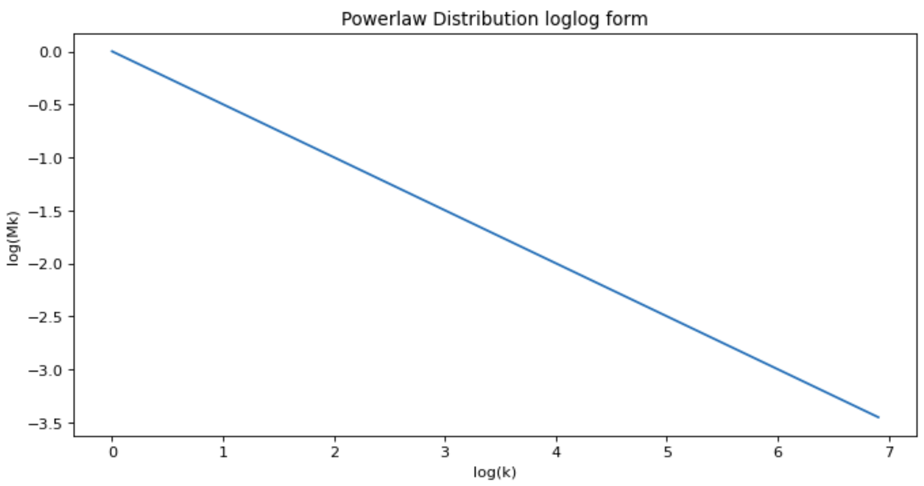
- *[1] Derek J. de Solla P.* ***Networks of scientific papers****, 1965, doi:10.1126/science.149.3683.510.*
- *[2] Michalis F., Petros F., and Christos F.* ***On power-law relationships of the Internet topolog****y, 1999, doi:10.1145/316188.316229.*
- *[3] Bollobás B. and Riordan O.* ***Handbook of Graphs and Networks: From the Genome to the Internet,*** *2003. Pages 1–34.*

POWERLAW IN REAL WORLD NETWORKS

- Is often believed that most real-world networks are scale-free [1], [2], [3].

Powerlaw $M_k \sim Ck^{-\lambda}$





References:

- *[1] Derek J. de Solla P. **Networks of scientific papers**, 1965, doi:10.1126/science.149.3683.510.*
- *[2] Michalis F., Petros F., and Christos F. **On power-law relationships of the Internet topolog**y, 1999, doi:10.1145/316188.316229.*
- *[3] Bollobás B. and Riordan O**. Handbook of Graphs and Networks: From the Genome to the Internet,** 2003. Pages 1–34.*

POWERLAW IN REAL WORLD NETWORKS

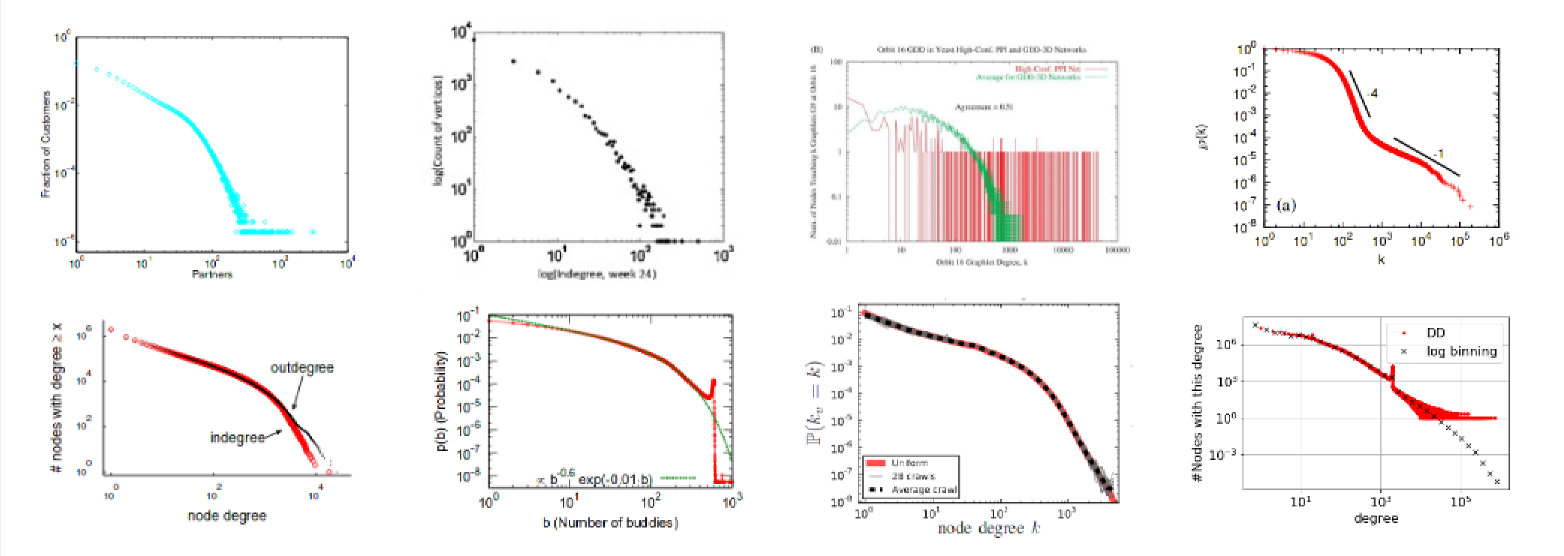
- Ubiquitousness of power law has been questioned [4].

References:

- *[4] Broido A. D.and A. Clauset, **"Scale-free networks are rare",** 2019, doi: 10.1038/s41467-019-08746-5.*

POWERLAW IN REAL WORLD NETWORKS

- Ubiquitousness of power law has been questioned [4].
- Investigations shown that **scale-free networks** are not so widespread as thought, it turns out that many of them follow a **power-law distribution with an exponential cut-off [5], [6].**



(biology, online market, social networks, ...)

References:

- *[4] Broido A. D.and A. Clauset, **“Scale-free networks are rare”,** 2019, doi: 10.1038/s41467-019-08746-5.*
- *[5] Newman M. E. J. **Coauthorship networks and patterns of scientific collaboration,** 2004, doi:10.1073/pnas.0307545100.*
- *[6] Newman M. E. J. **Clustering and preferential attachment in growing networks,** 2001, doi:10.1103/PhysRev*

POWERLAW IN REAL WORLD NETWORKS

- Ubiquitousness of power law has been questioned [4].
- Investigations shown that **scale-free networks** are not so widespread as thought, it turns out that many of them follow a **power-law distribution with an exponential cut-off [5], [6].**

Powerlaw with exponential cut-off $M_k \sim Ck^{-\lambda}\gamma^{-k} \mid 0 \leq \gamma < 1$

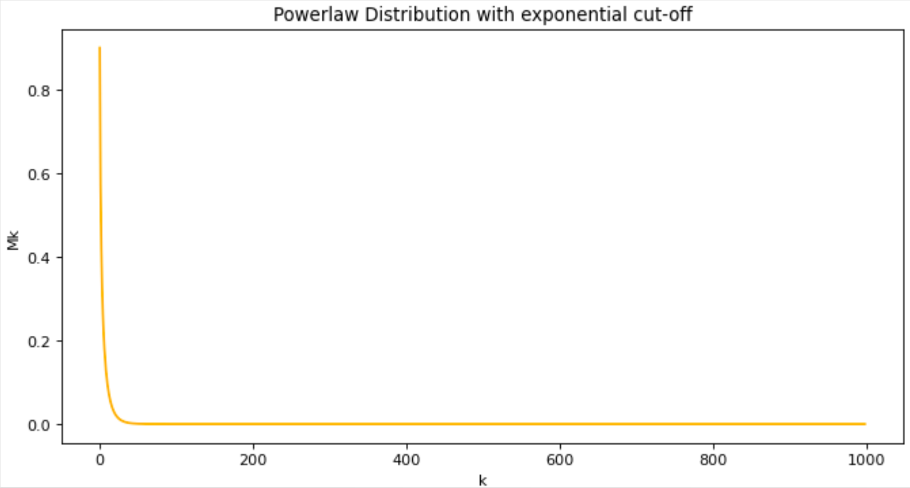
References:

- *[4] Broido A. D.and A. Clauset, **"Scale-free networks are rare",** 2019, doi: 10.1038/s41467-019-08746-5.*
- *[5] Newman M. E. J. **Coauthorship networks and patterns of scientific collaboration,** 2004, doi:10.1073/pnas.0307545100.*
- *[6] Newman M. E. J. **Clustering and preferential attachment in growing networks,** 2001, doi:10.1103/PhysRev*

POWERLAW IN REAL WORLD NETWORKS

- Ubiquitousness of power law has been questioned [4].
- Investigations shown that **scale-free networks** are not so widespread as thought, it turns out that many of them follow a **power-law distribution with an exponential cut-off [5], [6].**

Powerlaw with exponential cut-off $M_k \sim Ck^{-\lambda}\gamma^{-k}$ | $0 \leq \gamma < 1$



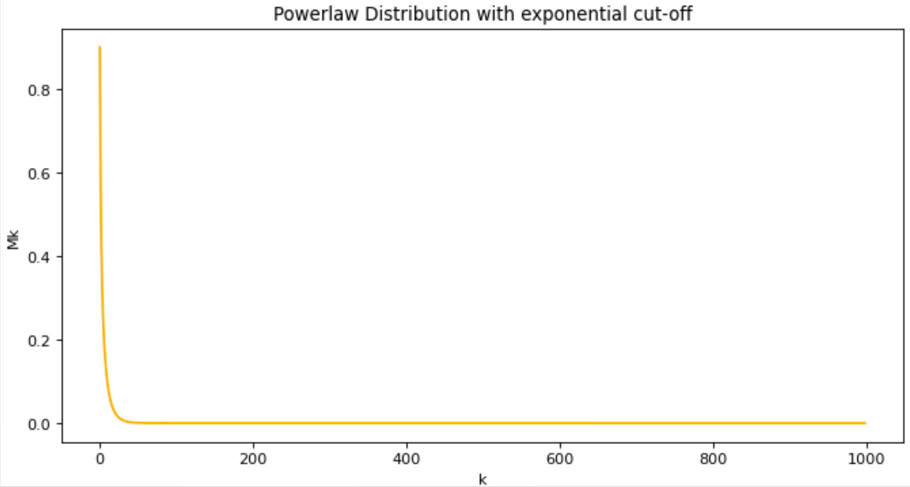
References:

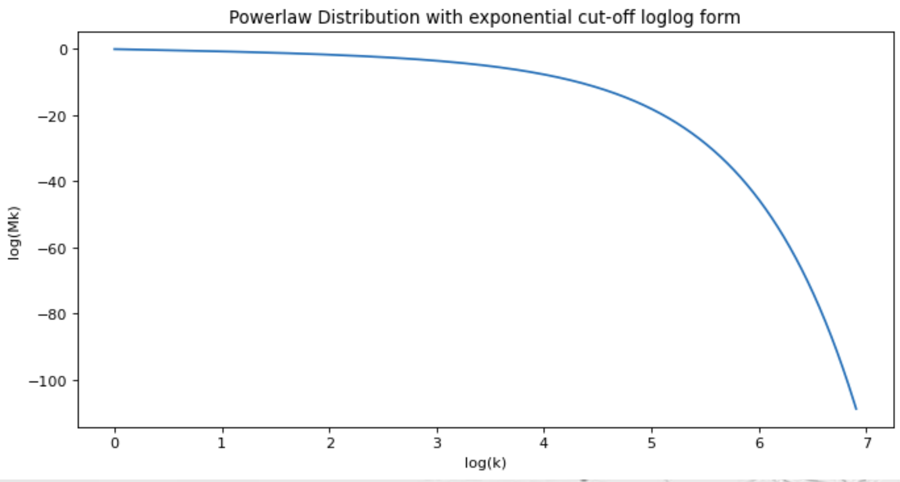
- *[4] Broido A. D.and A. Clauset, **"Scale-free networks are rare",** 2019, doi: 10.1038/s41467-019-08746-5.*
- *[5] Newman M. E. J. **Coauthorship networks and patterns of scientific collaboration,** 2004, doi:10.1073/pnas.0307545100.*
- *[6] Newman M. E. J. **Clustering and preferential attachment in growing networks,** 2001, doi:10.1103/PhysRev*

POWERLAW IN REAL WORLD NETWORKS

- Ubiquitousness of power law has been questioned [4].
- Investigations shown that **scale-free networks** are not so widespread as thought, it turns out that many of them follow a **power-law distribution with an exponential cut-off [5], [6].**

Powerlaw with exponential cut-off $M_k \sim Ck^{-\lambda}\gamma^{-k} \mid 0 \leq \gamma < 1$

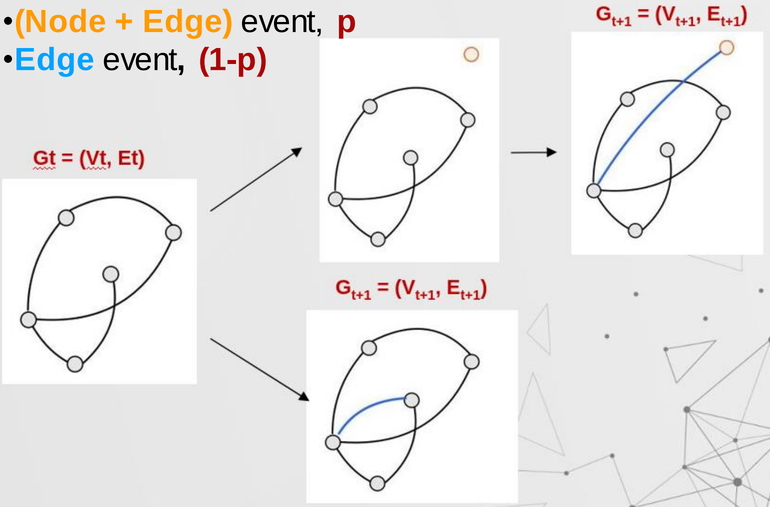




References:
- *[4] Broido A. D.and A. Clauset, **"Scale-free networks are rare",** 2019, doi: 10.1038/s41467-019-08746-5.*
- *[5] Newman M. E. J. **Coauthorship networks and patterns of scientific collaboration,** 2004, doi:10.1073/pnas.0307545100.*
- *[6] Newman M. E. J. **Clustering and preferential attachment in growing networks,** 2001, doi:10.1103/PhysRev*

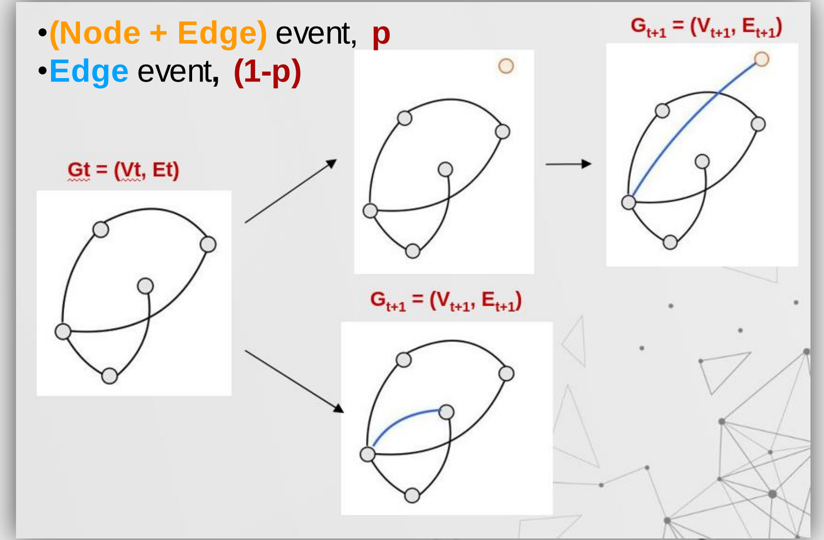
THEORETICAL MODEL

- (Node + Edge) event, p
- Edge event, (1-p)



THEORETICAL MODEL

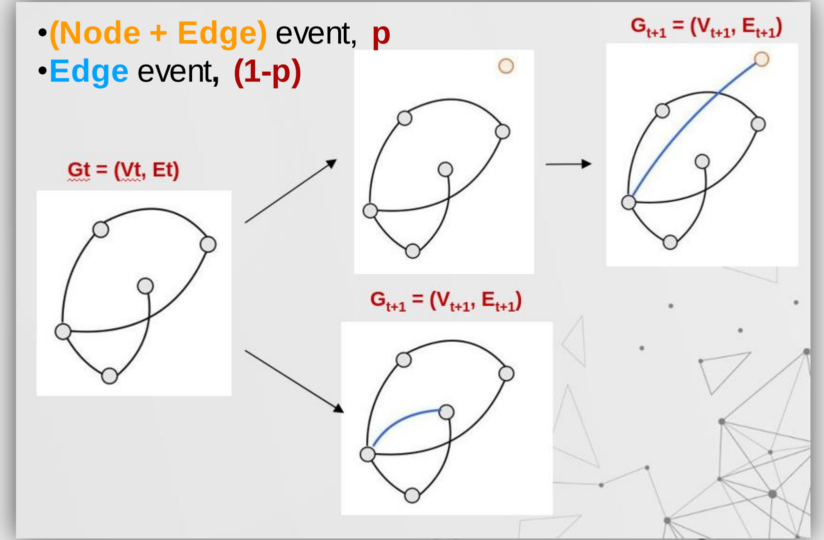
$\Pr[v \text{ is chosen at time step t}] =$



THEORETICAL MODEL

- The preferrential attachment is dictated by the so called **attachment function** $f(d_v(t)) = d_v(t)^{\gamma}$

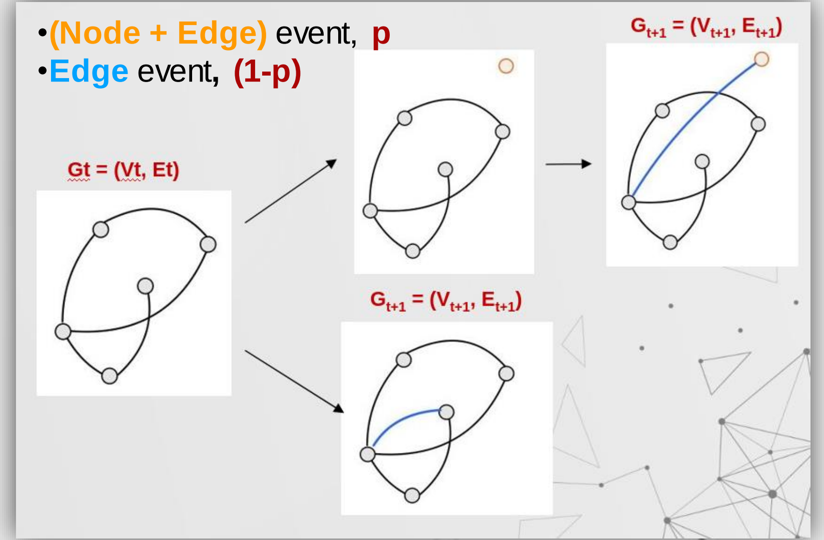
$$\Pr[v \text{ is chosen at time step t}] =$$



THEORETICAL MODEL

- The preferrential attachment is dictated by the so called **attachment function** $\mathbf{f(d_v(t)) = d_v(t)^\gamma}$

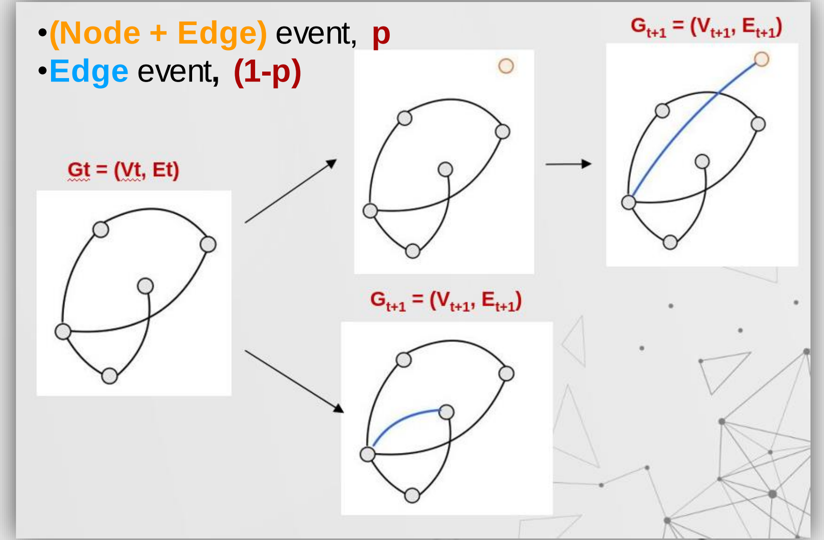
$$\Pr[v \text{ is chosen at time step t}] = \frac{f\left(\mathrm{d}_t(v)\right)}{\sum_{w \in V_t} f\left(\mathrm{d}_t(w)\right)}$$



THEORETICAL MODEL

- The preferrential attachment is dictated by the so called **attachment function** $f(d_v(t)) = d_v(t)^{\gamma}$

$$\Pr[v \text{ is chosen at time step t}] = \frac{f\left(\mathrm{d}_t(v)\right)}{\sum_{w \in V_t} f\left(\mathrm{d}_t(w)\right)}$$

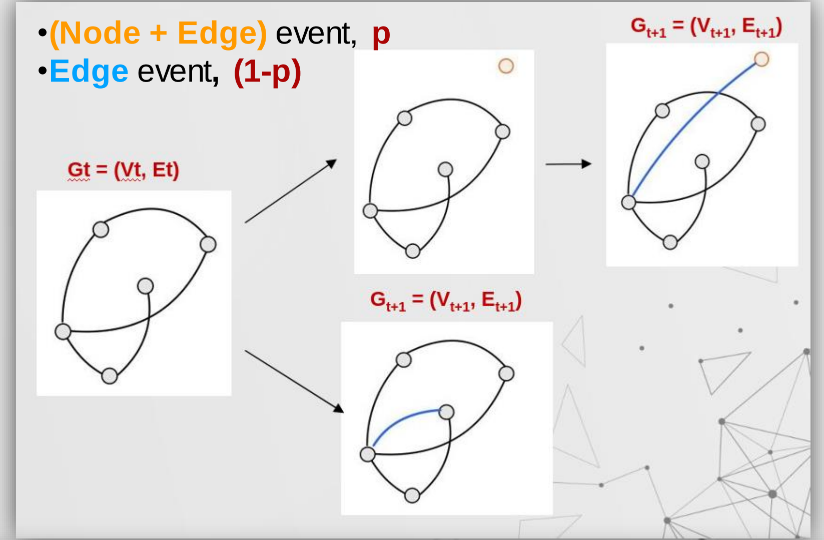


Vertex trajectory:

THEORETICAL MODEL

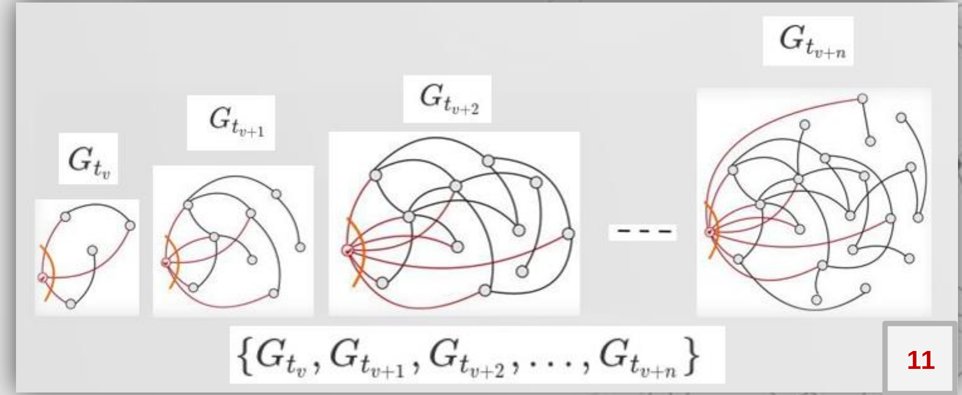
- The preferrential attachment is dictated by the so called **attachment function** $\mathbf{f(d_v(t)) = d_v(t)^{\gamma}}$

$$\Pr[v \text{ is chosen at time step t}] = \frac{f\left(\mathrm{d}_t(v)\right)}{\sum_{w \in V_t} f\left(\mathrm{d}_t(w)\right)}$$



Vertex trajectory:

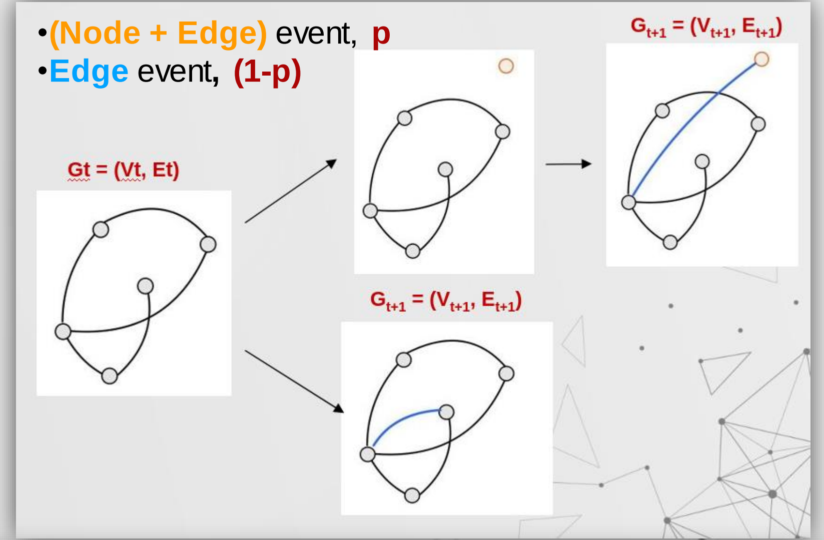
- Let t_v time step in which **v** appears in the graph;



THEORETICAL MODEL

- The preferrential attachment is dictated by the so called **attachment function** $f(d_v(t)) = d_v(t)^\gamma$

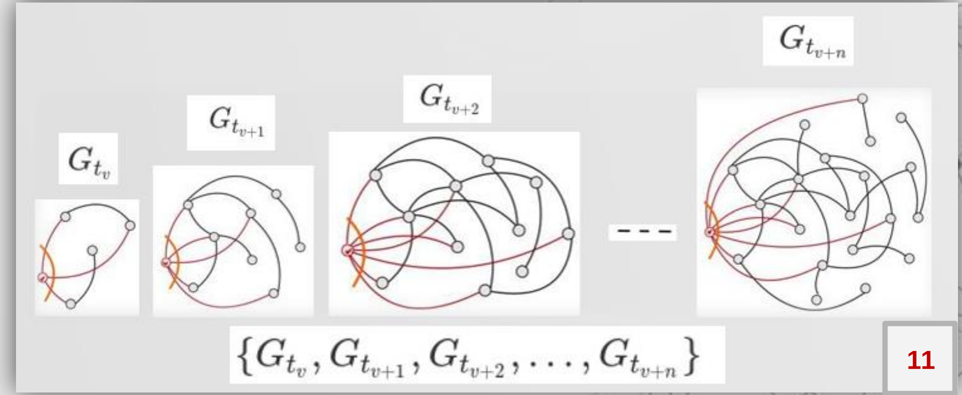
$$\Pr[v \text{ is chosen at time step t}] = \frac{f\left(\mathrm{d}_t(v)\right)}{\sum_{w \in V_t} f\left(\mathrm{d}_t(w)\right)}$$



Vertex trajectory:

- Let t_v time step in which **v** appears in the graph;
- The vertex trajectory $d_v(t)$ is the sequence:

$$\{d_v(t_v), d_v(t_{v+1}), d_v(t_{v+2}), \ldots, d_v(t_{v+n})\}$$



MATCHING MODEL

Barabasi Albert model [7]

Given:

- **P = 1**, so only **node+edge** event
- **γ = 1,** so $\mathbf{f(d_v(t)) = d_v(t)}$

References:

- *[7] Barabasi A. and Albert R. **Emergence of Scaling in Random Networks.** 1999, Pages 509–12, doi: 10.1126/science.286.5439.509 5*

MATCHING MODEL

Barabasi Albert model [7]

Given:

- $P = 1$, so only **node+edge** event
- $\gamma = 1$, so $\mathbf{f(d_v(t)) = d_v(t)}$

Follows:

- Power law degree distribution $M_k \sim Ck^{-\lambda}$
- Vertext trajectory $d_v(t) = (t/t_v)^{1/2}$

References:

- *[7] Barabasi A. and Albert R.* ***Emergence of Scaling in Random Networks.*** *1999, Pages 509–12, doi: 10.1126/science.286.5439.509 5*

MATCHING MODEL

Barabasi Albert model [7]

Given:

- $P = 1$, so only **node+edge** event
- $\gamma = 1$, so $\mathbf{f(d_v(t)) = d_v(t)}$

Follows:

- Power law degree distribution $M_k \sim Ck^{-\lambda}$
- Vertext trajectory $d_v(t) = (t/t_v)^{1/2}$

Chung-Lu model [8]

Given:

- General P and $0 \leq \gamma < 1$
- Generalized sublinear $\mathbf{f(d_v(t)) = d_v(t)}^{\gamma}$

References:

- *[7] Barabasi A. and Albert R. **Emergence of Scaling in Random Networks.** 1999, Pages 509–12, doi: 10.1126/science.286.5439.509.5*
- *[8] Giroire F., Nisse N., Sulkowska M., **Study of a degree distribution and a vertex trajectory in the Chung-Lu model with a generalized attachment function,** 2022*
- *[10] Chung F. and Lu L., **Complex Graphs and Networks**, 2006.*

MATCHING MODEL

Barabasi Albert model [7]

Given:
- $P = 1$, so only **node+edge** event
- $\gamma = 1$, so $\mathbf{f(d_v(t)) = d_v(t)}$

Follows:
- Power law degree distribution $M_k \sim Ck^{-\lambda}$
- Vertext trajectory $d_v(t) = (t/t_v)^{1/2}$

Chung-Lu model [8]

Given:
- General P and $0 \leq \gamma < 1$
- Generalized sublinear $\mathbf{f(d_v(t)) = d_v(t)^{\gamma}}$

Follows :
- **stretched exponential** degree distribution M_k

References:
- *[7] Barabasi A. and Albert R. **Emergence of Scaling in Random Networks.** 1999, Pages 509–12, doi: 10.1126/science.286.5439.509.5*
- *[8] Giroire F., Nisse N., Sulkowska M., **Study of a degree distribution and a vertex trajectory in the Chung-Lu model with a generalized attachment function,** 2022*
- *[10] Chung F. and Lu L., **Complex Graphs and Networks**, 2006.*

MATCHING MODEL

Barabasi Albert model [7]

Given:
- $P = 1$, so only **node+edge** event
- $\gamma = 1$, so $\mathbf{f(d_v(t)) = d_v(t)}$

Follows:
- Power law degree distribution $M_k \sim Ck^{-\lambda}$
- Vertext trajectory $d_v(t) = (t/t_v)^{1/2}$

Chung-Lu model [8]

Given:
- General P and $0 \leq \gamma < 1$
- Generalized sublinear $\mathbf{f(d_v(t)) = d_v(t)^{\gamma}}$

Follows :
- **stretched exponential** degree distribution M_k

$$M_k \sim \alpha \cdot k^{-\gamma} \cdot \exp\left\{-\frac{\alpha}{1-\gamma}k^{1-\gamma}\right\}$$

References:

- *[7] Barabasi A. and Albert R.* ***Emergence of Scaling in Random Networks.*** *1999, Pages 509–12, doi: 10.1126/science.286.5439.509 5*
- *[8] Giroire F., Nisse N., Sulkowska M.,* ***Study of a degree distribution and a vertex trajectory in the Chung-Lu model with a generalized attachment function,*** *2022*
- *[10] Chung F. and Lu L.,* ***Complex Graphs and Networks****, 2006.*

MATCHING MODEL

Barabasi Albert model [7]

Given:
- $P = 1$, so only **node+edge** event
- $\gamma = 1$, so $\mathbf{f(d_v(t)) = d_v(t)}$

Follows:
- Power law degree distribution $M_k \sim Ck^{-\lambda}$
- Vertext trajectory $d_v(t) = (t/t_v)^{1/2}$

Chung-Lu model [8]

Given:
- General P and $0 \leq \gamma < 1$
- Generalized sublinear $\mathbf{f(d_v(t)) = d_v(t)^{\gamma}}$

Follows :
- **stretched exponential** degree distribution M_k
- logarithmic vertex trajectory $d_v(t)$

$$M_k \sim \alpha \cdot k^{-\gamma} \cdot \exp\left\{-\frac{\alpha}{1-\gamma} k^{1-\gamma}\right\}$$

References:
- *[7] Barabasi A. and Albert R.* ***Emergence of Scaling in Random Networks.*** *1999, Pages 509–12, doi: 10.1126/science.286.5439.509 5*
- *[8] Giroire F., Nisse N., Sulkowska M.,* ***Study of a degree distribution and a vertex trajectory in the Chung-Lu model with a generalized attachment function,*** *2022*
- *[10] Chung F. and Lu L.,* ***Complex Graphs and Networks****, 2006.*

MATCHING MODEL

Barabasi Albert model [7]

Given:
- $P = 1$, so only **node+edge** event
- $\gamma = 1$, so $\mathbf{f(d_v(t)) = d_v(t)}$

Follows:
- Power law degree distribution $M_k \sim Ck^{-\lambda}$
- Vertext trajectory $d_v(t) = (t/t_v)^{1/2}$

Chung-Lu model [8]

Given:
- General P and $0 \leq \gamma < 1$
- Generalized sublinear $\mathbf{f(d_v(t)) = d_v(t)^{\gamma}}$

Follows :
- **stretched exponential** degree distribution M_k
- logarithmic vertex trajectory $d_v(t)$

$$M_k \sim \alpha \cdot k^{-\gamma} \cdot \exp\left\{-\frac{\alpha}{1-\gamma}k^{1-\gamma}\right\}$$

$$d_v(t) = \left(\frac{1-\gamma}{\alpha}\ln(t/t_v) + 1\right)^{1/(1-\gamma)}$$

References:
- *[7] Barabasi A. and Albert R. **Emergence of Scaling in Random Networks.** 1999, Pages 509–12, doi: 10.1126/science.286.5439.509 5*
- *[8] Giroire F., Nisse N., Sulkowska M., **Study of a degree distribution and a vertex trajectory in the Chung-Lu model with a generalized attachment function,** 2022*
- *[10] Chung F. and Lu L., **Complex Graphs and Networks**, 2006.*

03

FACED PROBLEMS

(1) Which researchers should be taken into account?
(2) How to interpret a time step?

(1) Which researchers should be taken into account?

Problem

We faced the following issues trying to define **what an active author is**:

(1) Which researchers should be taken into account?

Problem

We faced the following issues trying to define **what an active author is**:

1. Authors with **low publication rate**, too many years of inactivity;.

(1) Which researchers should be taken into account?

Problem

We faced the following issues trying to define **what an active author is**:

1. Authors with **low publication rate**, too many years of inactivity;
2. Authors with **few years of research activity**, like students who stopped publishing after their PhD;

(1) Which researchers should be taken into account?

Problem

We faced the following issues trying to define **what an active author is**:

1. Authors with **low publication rate**, too many years of inactivity;
2. Authors with **few years of research activity**, like students who stopped publishing after their PhD;
3. Authors with **few publications**.

(1) Which researchers should be taken into account?

Problems

We faced the following issues trying to define **what an active author is**:

1. Authors with **low publication rate**, too many years of inactivity;
2. Authors with **few years of research activity**, like students who stopped publishing after their PhD;
3. Authors with **few publications**.

Solutions:

We defined active authors by the following metrics:

(1) Which researchers should be taken into account?

Problems

We faced the following issues trying to define **what an active author is**:

1. Authors with **low publication rate**, too many years of inactivity;
2. Authors with **few years of research activity**, like students who stopped publishing after their PhD;
3. Authors with **few publications**.

Solutions:

We defined active authors by the following metrics:

1. **Hole size**: # consecutive years without publishing;

(1) Which researchers should be taken into account?

Problems

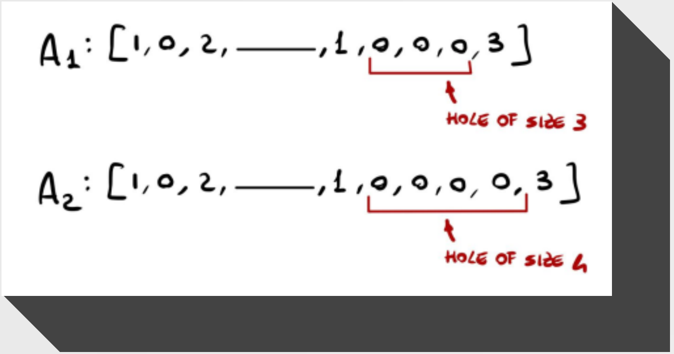
We faced the following issues trying to define **what an active author is**:

1. Authors with **low publication rate**, too many years of inactivity;
2. Authors with **few years of research activity**, like students who stopped publishing after their PhD;
3. Authors with **few publications**.

Solutions:

We defined active authors by the following metrics:

1. **Hole size**: # consecutive years without publishing;



(1) Which researchers should be taken into account?

Problems

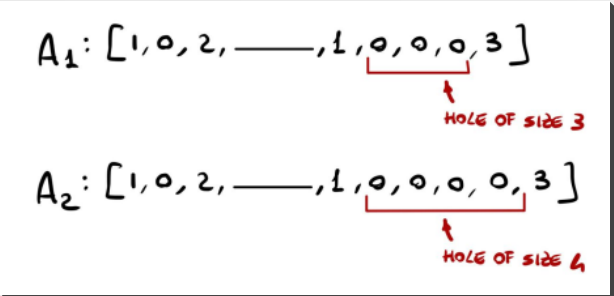
We faced the following issues trying to define **what an active author is**:

1. Authors with **low publication rate**, too many years of inactivity;
2. Authors with **few years of research activity**, like students who stopped publishing after their PhD;
3. Authors with **few publications**.

Solutions:

We defined active authors by the following metrics:

1. **Hole size**: #consecutive years without publishing;
2. **Activity period:** #years between 1st and last publication;



(1) Which researchers should be taken into account?

Problems

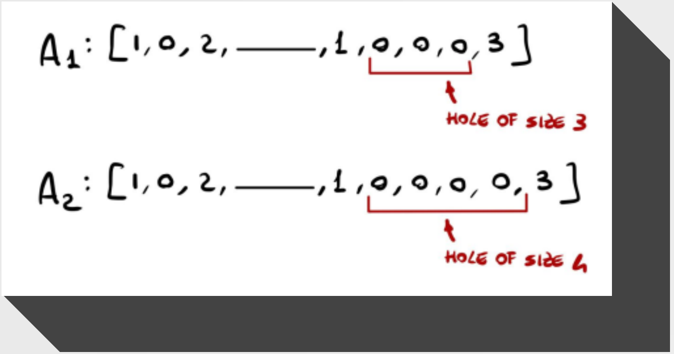
We faced the following issues trying to define **what an active author is**:

1. Authors with **low publication rate**, too many years of inactivity;
2. Authors with **few years of research activity**, like students who stopped publishing after their PhD;
3. Authors with **few publications**.

Solutions:

We defined active authors by the following metrics:

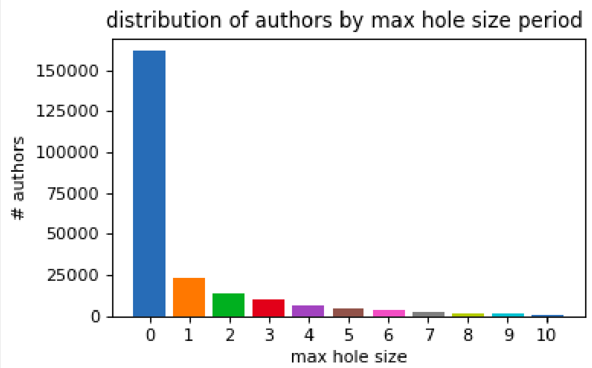
1. **Hole size**: #consecutive years without publishing;
2. **Activity period:** #years between 1st and last publication;
3. **Publication number**: total #publications.



(1) Which researchers should be taken into account?

Hole size

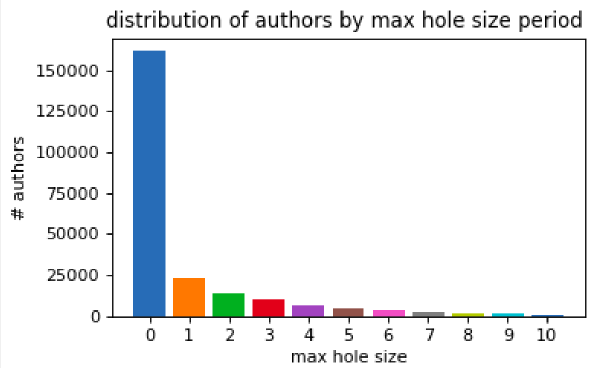
(# consecutive years without publishing)



(1) Which researchers should be taken into account?

Hole size

(# consecutive years without publishing)



To include researcher who take a sabathical year from teaching every seven, to do research

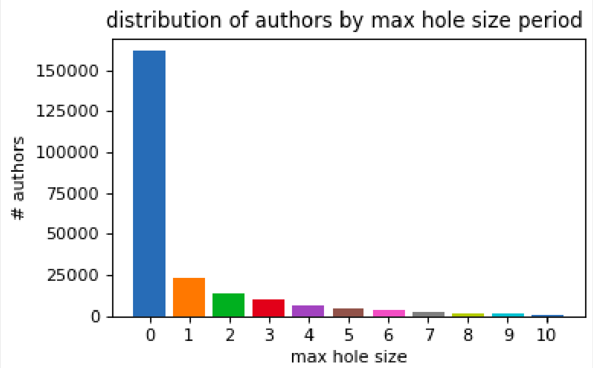
≤ 7

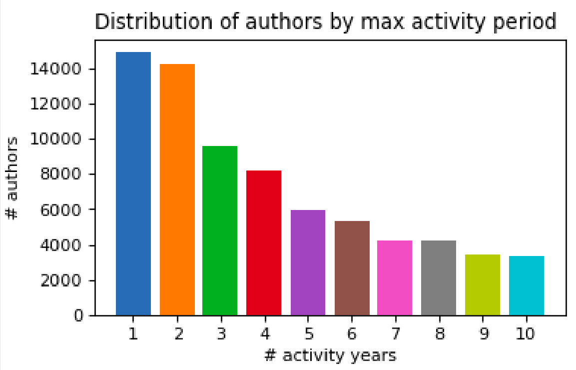
(1) Which researchers should be taken into account?

Hole size
(# consecutive years without publishing)

&

Activity period
(#years between 1st and last publication)





To include researcher who take a sabathical year from teaching every seven, to do research

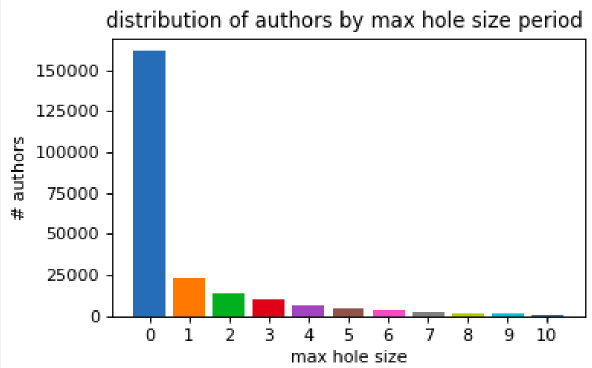
≤ 7

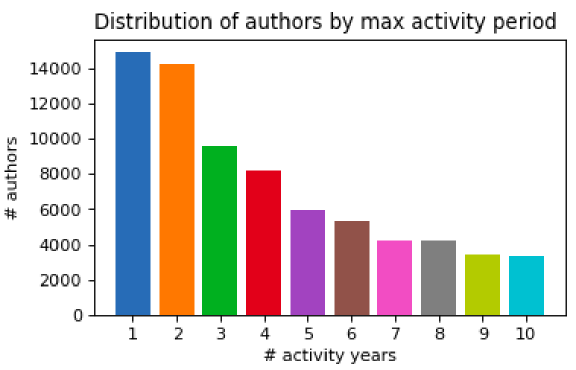
(1) Which researchers should be taken into account?

Hole size
(# consecutive years without publishing)

&

Activity period
(#years between 1st and last publication)





To include researcher who take a sabathical year from teaching every seven, to do research

≤ 7

To exlude PhD students

≥ 5

(1) Which researchers should be taken into account?

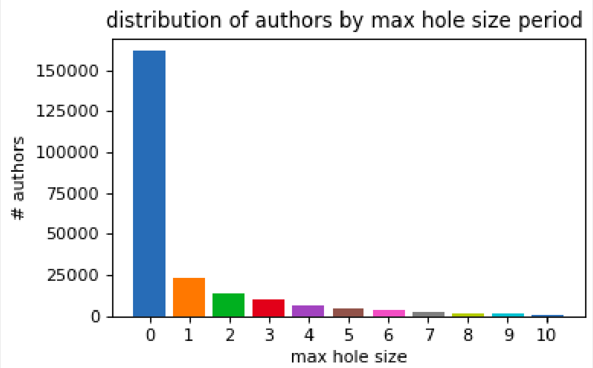
Hole size
(# consecutive years without publishing)

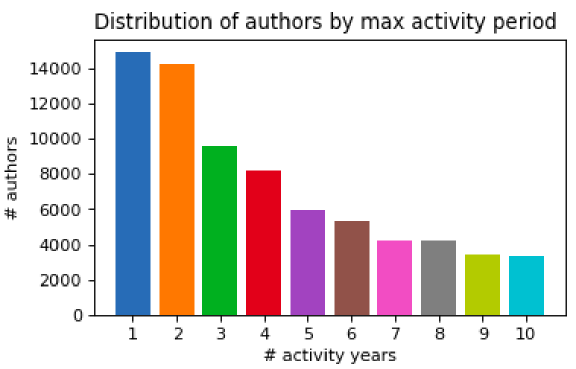
&

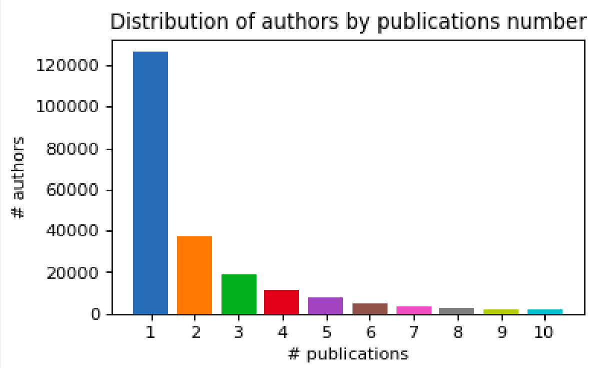
Activity period
(#years between 1st and last publication)

&

#publications







To include researcher who take a sabathical year from teaching every seven, to do research

To exlude PhD students

≤ 7

≥ 5

(1) Which researchers should be taken into account?

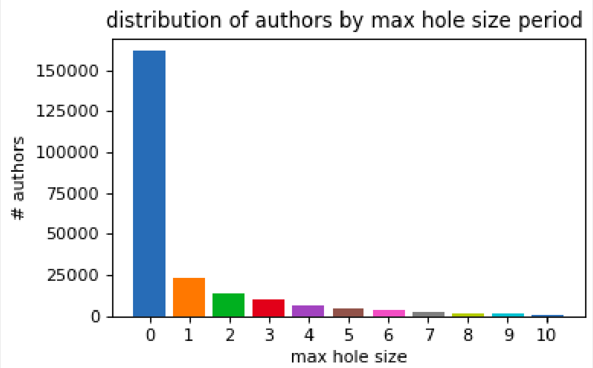
Hole size
(# consecutive years without publishing)

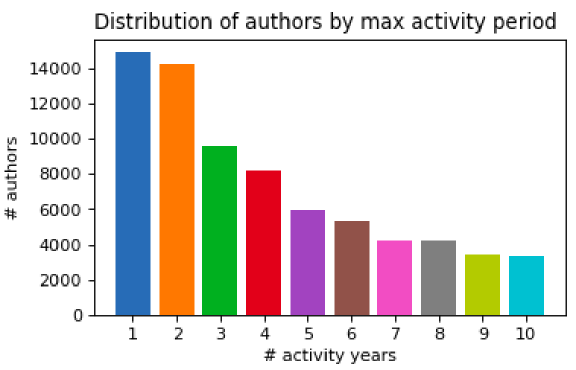
&

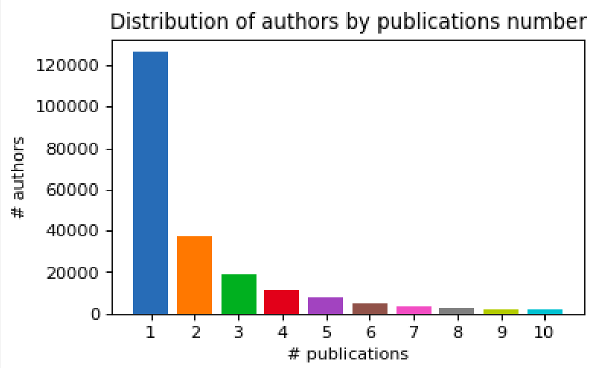
Activity period
(#years between 1st and last publication)

&

#publications







To include researcher who take a sabathical year from teaching every seven, to do research

≤ 7

To exlude PhD students

≥ 5

Good compromise between insight and #authors filtered

≥ 3

(1) Which researchers should be taken into account?

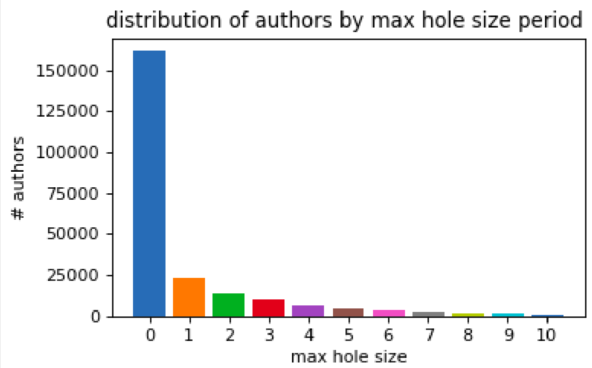
Hole size
(# consecutive years without publishing)

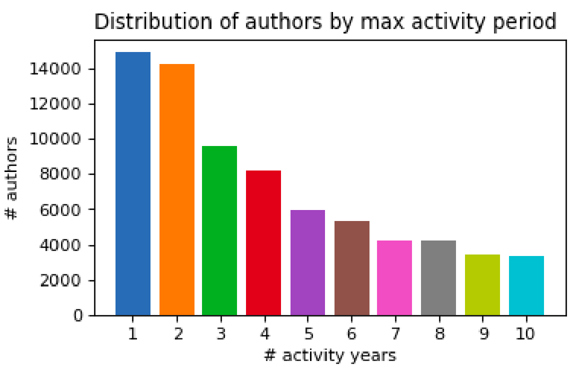
&

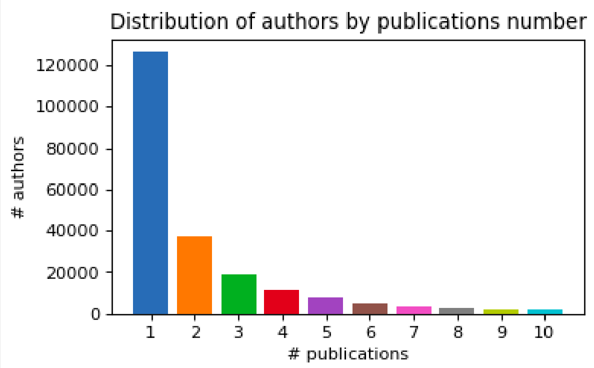
Activity period
(#years between 1st and last publication)

&

#publications







To include researcher who take a sabathical year from teaching every seven, to do research

≤ 7

To exlude PhD students

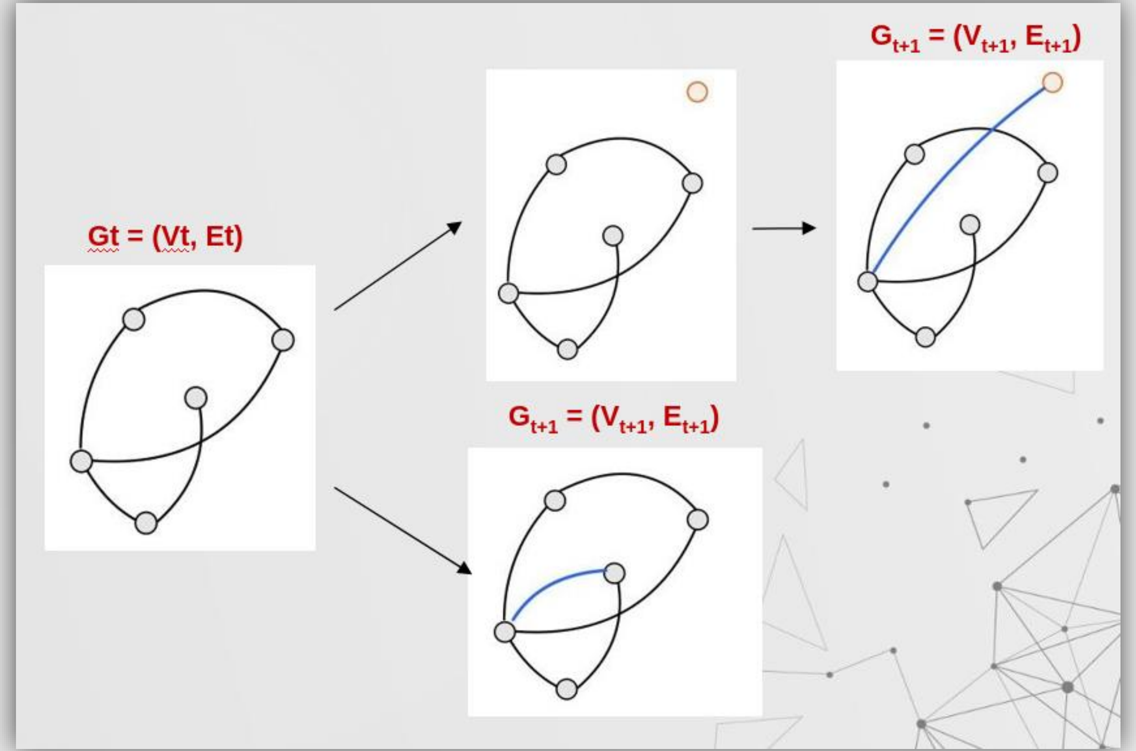
≥ 5

Good compromise between insight and #authors filtered

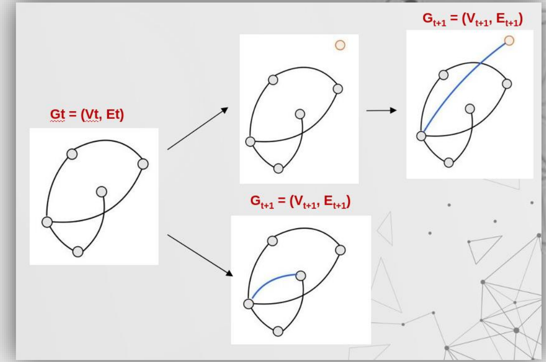
≥ 3

=> 16% of data left

(2) How to interpret a time step?



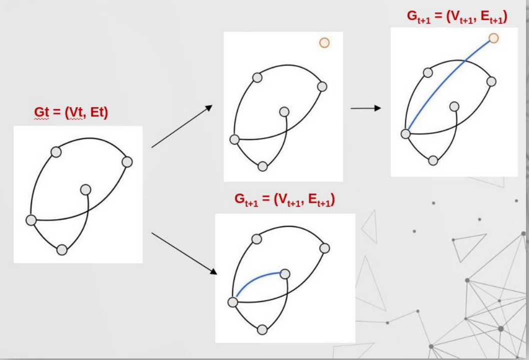
(2) How to interpret a time step?



- Time step definition:

  $G_t = (V_t, E_t)$ graph after event **t | t∈N** - #events occurred (**(Node + Edge)** **and** **Edge** event)

(2) How to interpret a time step?

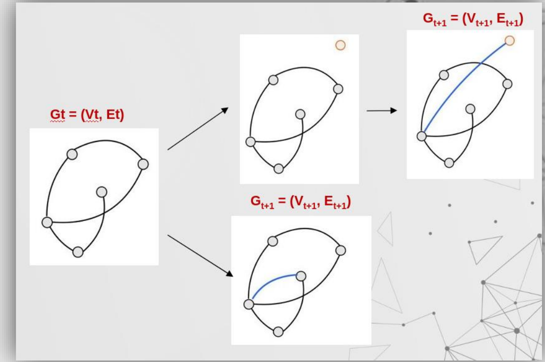


- Time step definition:

  $G_t = (V_t, E_t)$ graph after event **t | t∈N** - #events occurred (**(Node + Edge) and Edge** event)

- We have the **state of the graph for each year**, but our model not refer to the year but to the appearance of author or collaboration.

(2) How to interpret a time step?



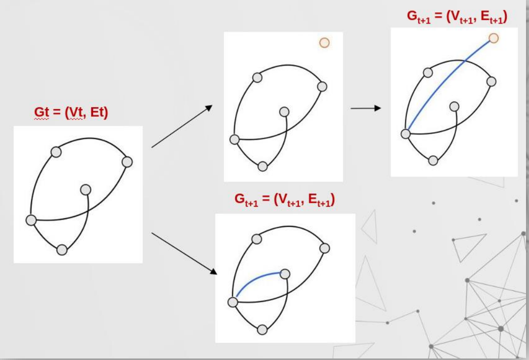
- Time step definition:

  $G_t = (V_t, E_t)$ graph after event **t | t∈N** – #events occurred (**(Node + Edge)** **and** **Edge** event)

- We have the **state of the graph for each year**, but our model not refer to the year but to the appearance of author or collaboration.

- Other metrics as event : the occurrence of a **new author**

(2) How to interpret a time step?

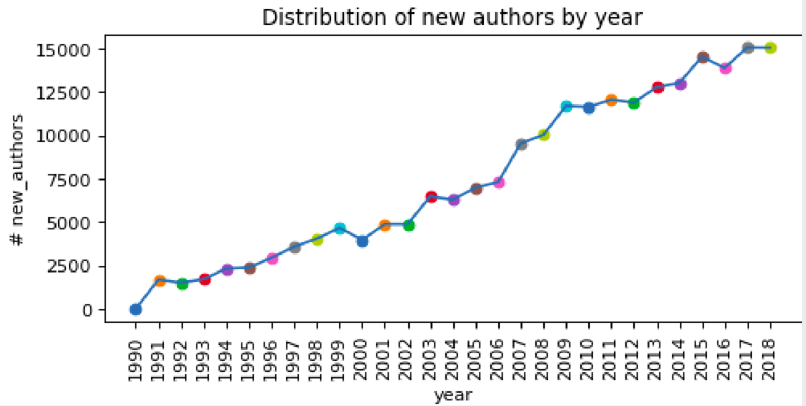


- Time step definition:

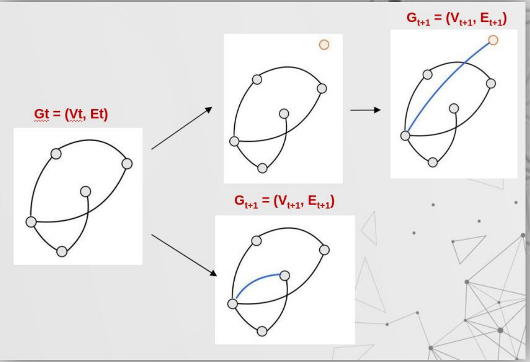
  $G_t = (V_t, E_t)$ graph after event **t | t∈N** - #events occurred (**(Node + Edge)** **and** **Edge** event)

- We have the **state of the graph for each year**, but our model not refer to the year but to the appearance of author or collaboration.

- Other metrics as event : the occurrence of a **new author**



(2) How to interpret a time step?

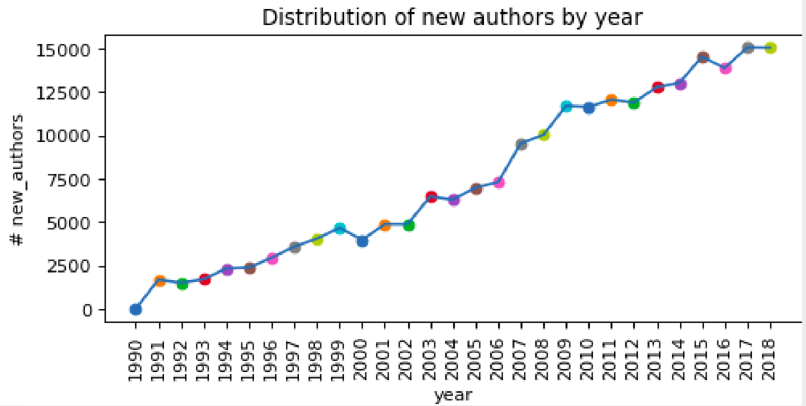


- Time step definition:

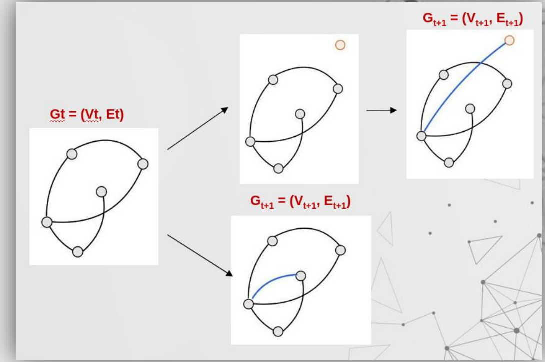
  $G_t = (V_t, E_t)$ graph after event **t | t∈N** – #events occurred (**(Node + Edge)** **and** **Edge** event)

- We have the **state of the graph for each year**, but our model not refer to the year but to the appearance of author or collaboration.

- Other metrics as event : the occurrence of a **new author** or a **new collaboration**.



(2) How to interpret a time step?

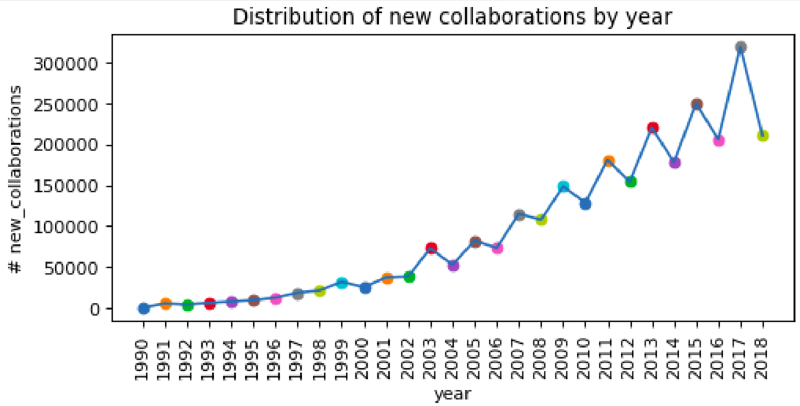


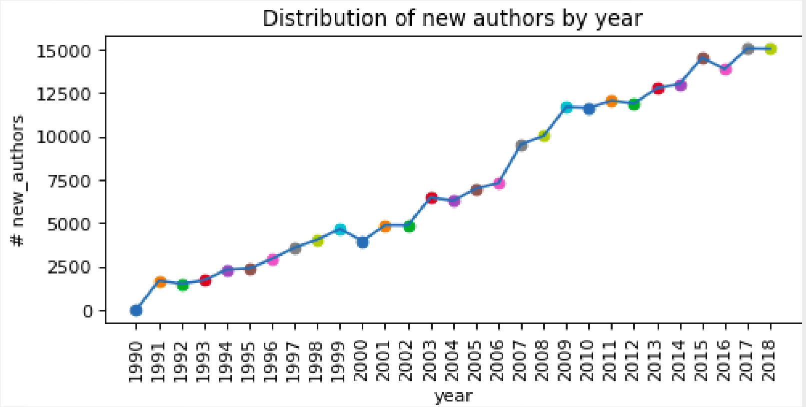
- Time step definition:

  $G_t = (V_t, E_t)$ graph after event **t | t∈N** – #events occurred (**(Node + Edge)** and **Edge** event)

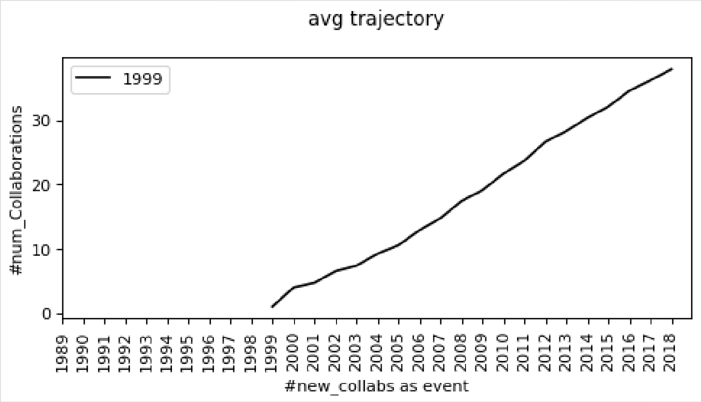
- We have the **state of the graph for each year**, but our model not refer to the year but to the appearance of author or collaboration.

- Other metrics as event : the occurrence of a **new author** or a **new collaboration**.





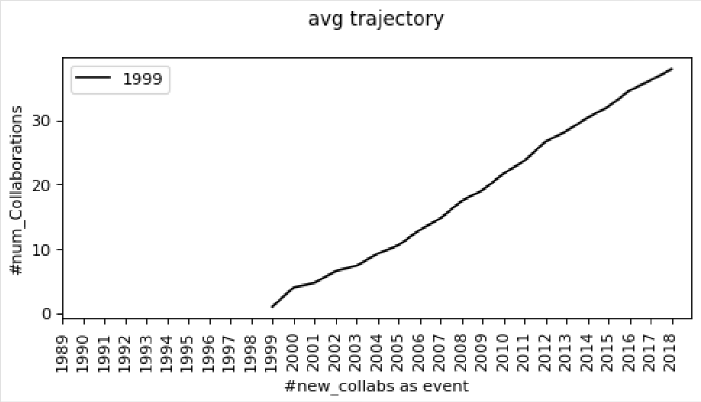
(2) How to interpret a time step?



(2) How to interpret a time step?

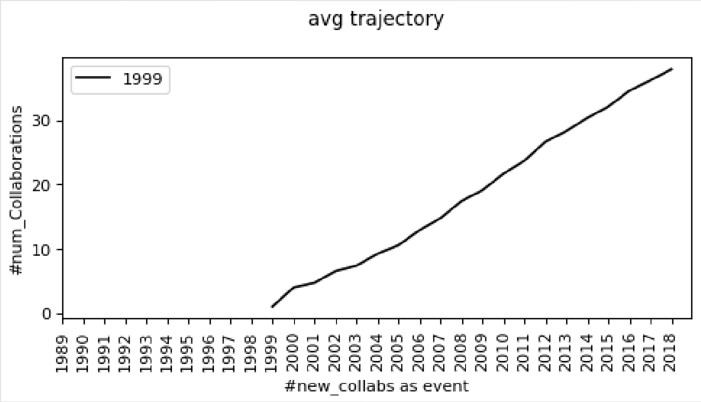
- Vertex trajectories are **stretched on the x axis**

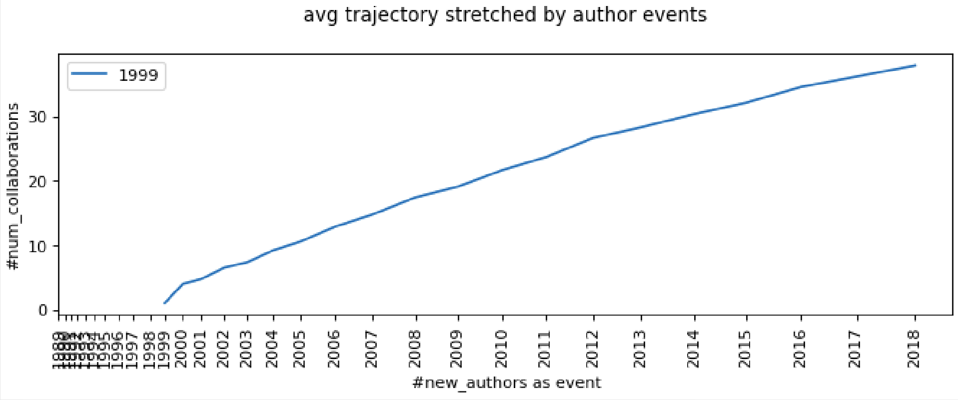
- Shown their **logarithmic** shape.

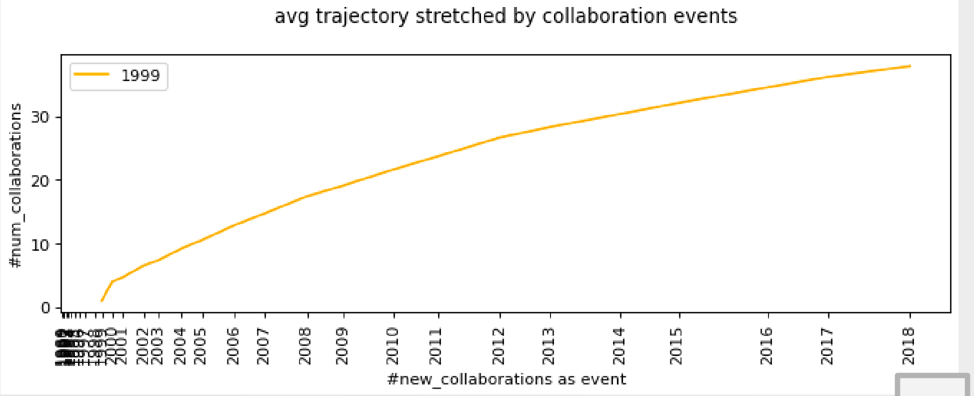


(2) How to interpret a time step?

- Vertex trajectories are **stretched on the x axis**
- Shown their **logarithmic** shape.







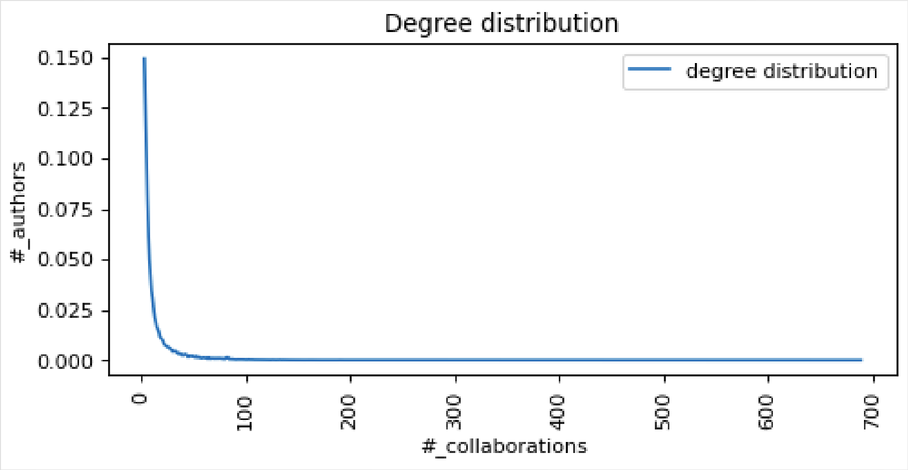
04

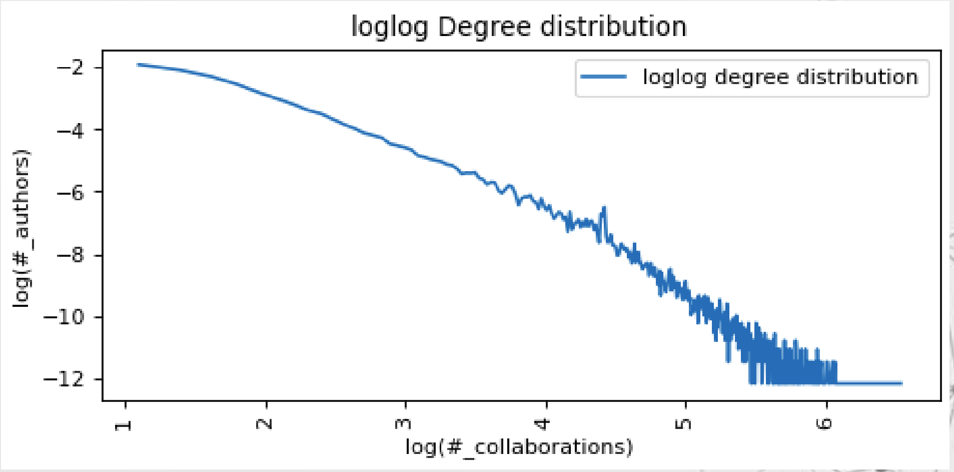
DEGREE DISTRIBUTION

Degree Distribution Retrieval
Degree Distribution Fitting

DEGREE DISTRIBUTION RETRIEVAL

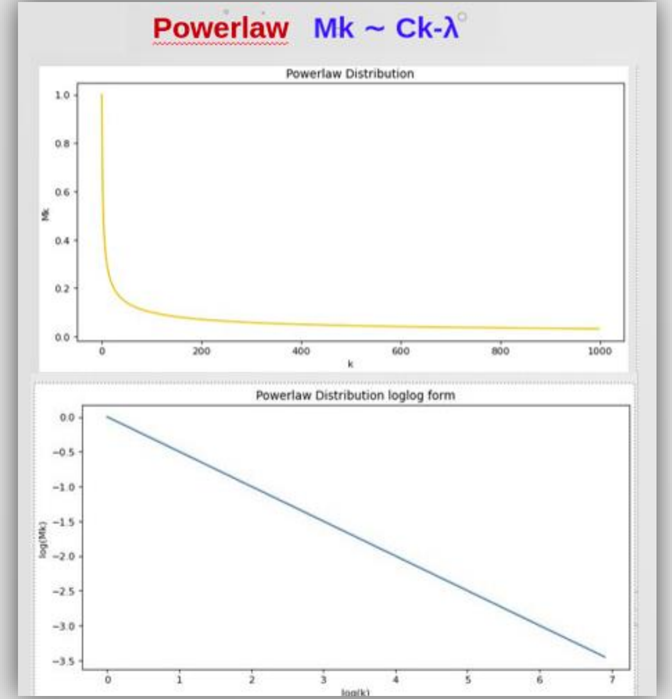
To acquire more knowledge from the character of the data the **degree distribution of active authors** is found, taking the **#authors** with a given **#collaborations.**





DEGREE DISTRIBUTION FITTING

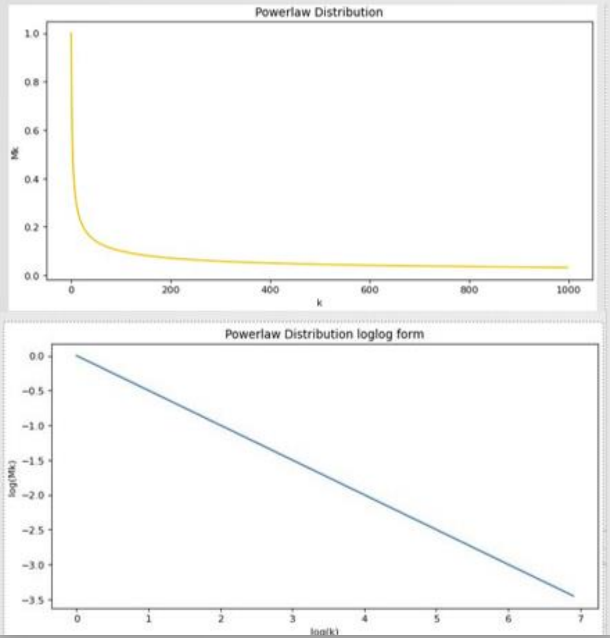
To acquire more knowledge from the character of the data the **degree distribution of active authors** is found, taking the **#authors** with a given **#collaborations.**



DEGREE DISTRIBUTION FITTING

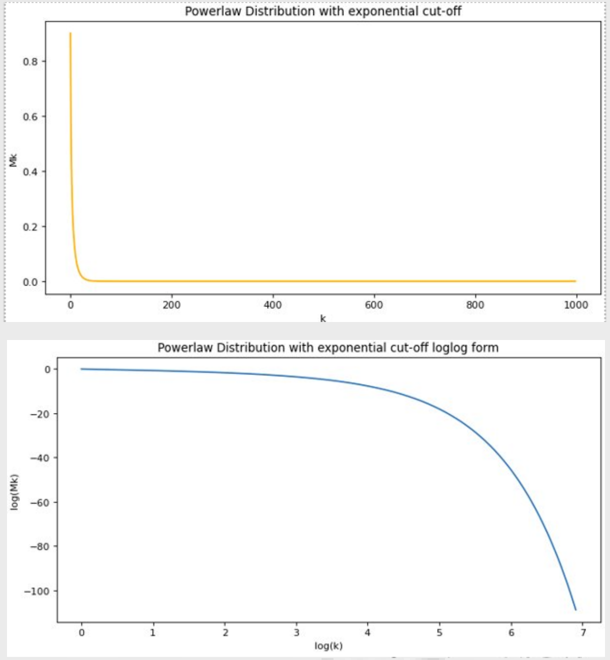
To acquire more knowledge from the character of the data the **degree distribution of active authors** is found, taking the **#authors** with a given **#collaborations.**

Powerlaw $Mk \sim Ck^{-\lambda}$



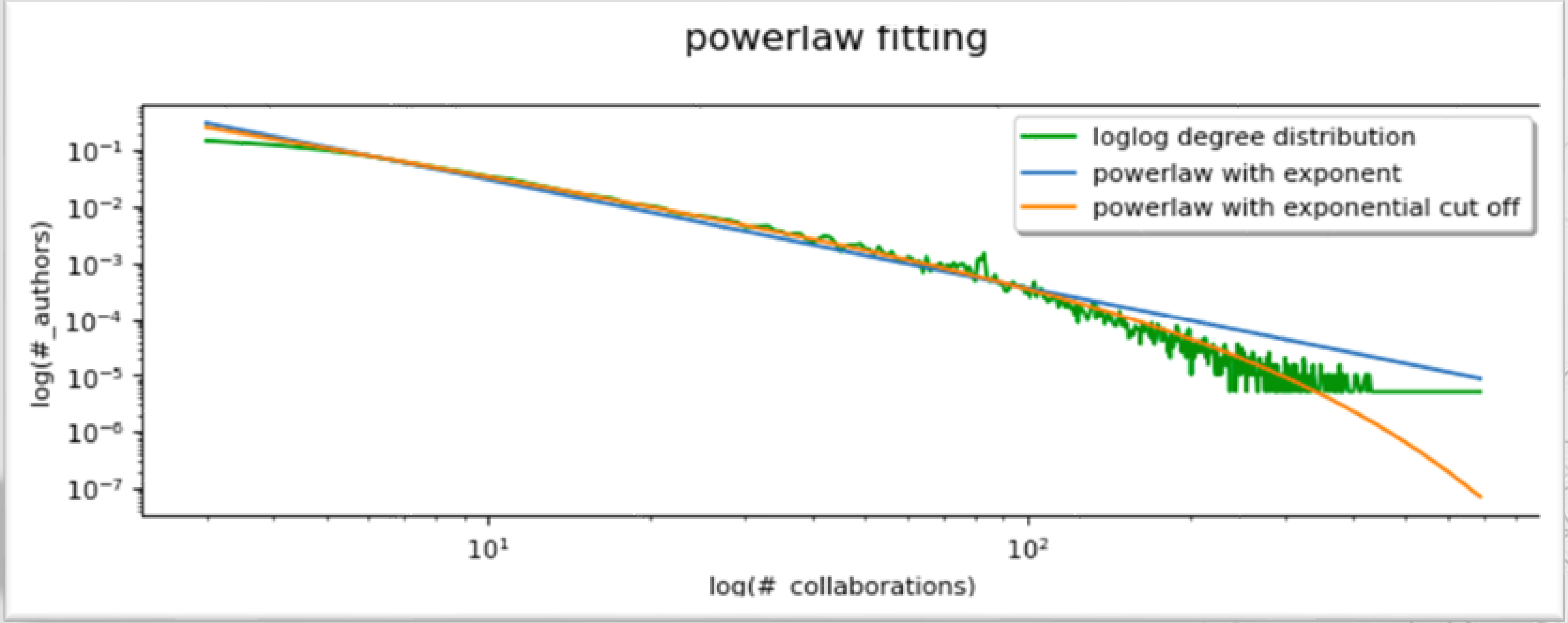
Powerlaw with exponential cut-off

$Mk \sim Ck^{-\lambda}\gamma^{-k} \mid 0 \leq \gamma < 1$



DEGREE DISTRIBUTION RETRIEVAL

To acquire more knowledge from the character of the data the **degree distribution of active authors** is found, taking the **#authors** with a given **#collaborations.**



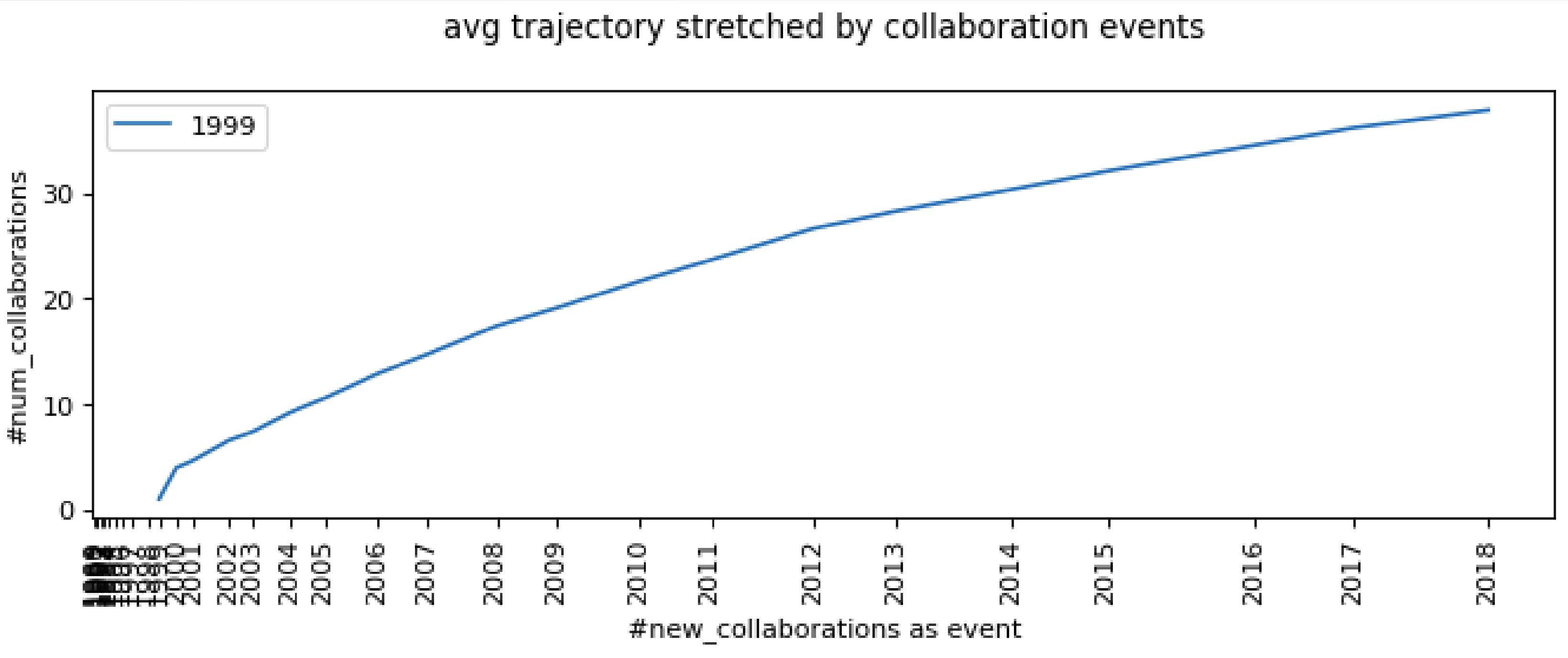
05

VERTEX TRAJECTORIES

Average Trajectories
Fitting Trajectories
General Fitting
Final Contraddiction

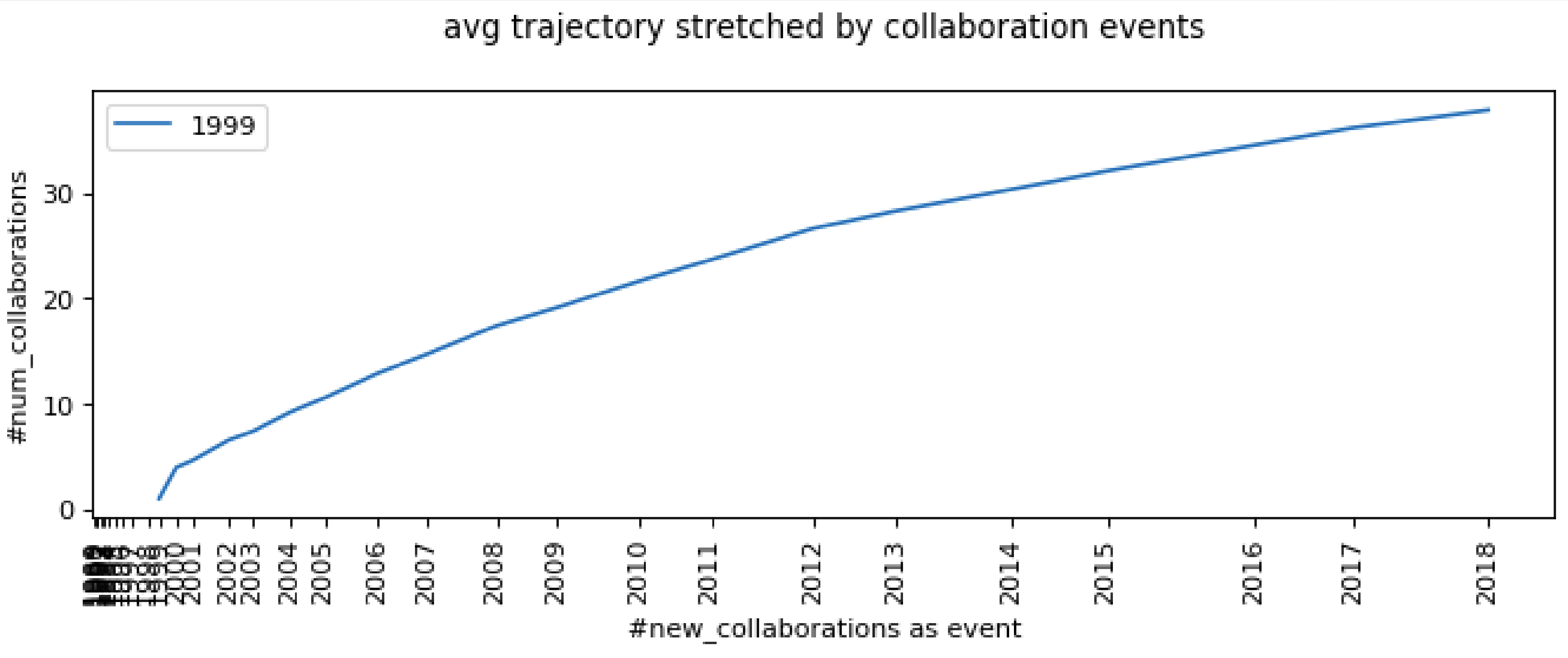
AVERAGE TRAJECTORIES

- Computed and plotted the **average trajectory of active authors** by starting year.



AVERAGE TRAJECTORIES

- Computed and plotted the **average trajectory of active authors** by starting year.
- The x-axis is stretched by the number of new collaborations as events.



FITTING TRAJECTORIES

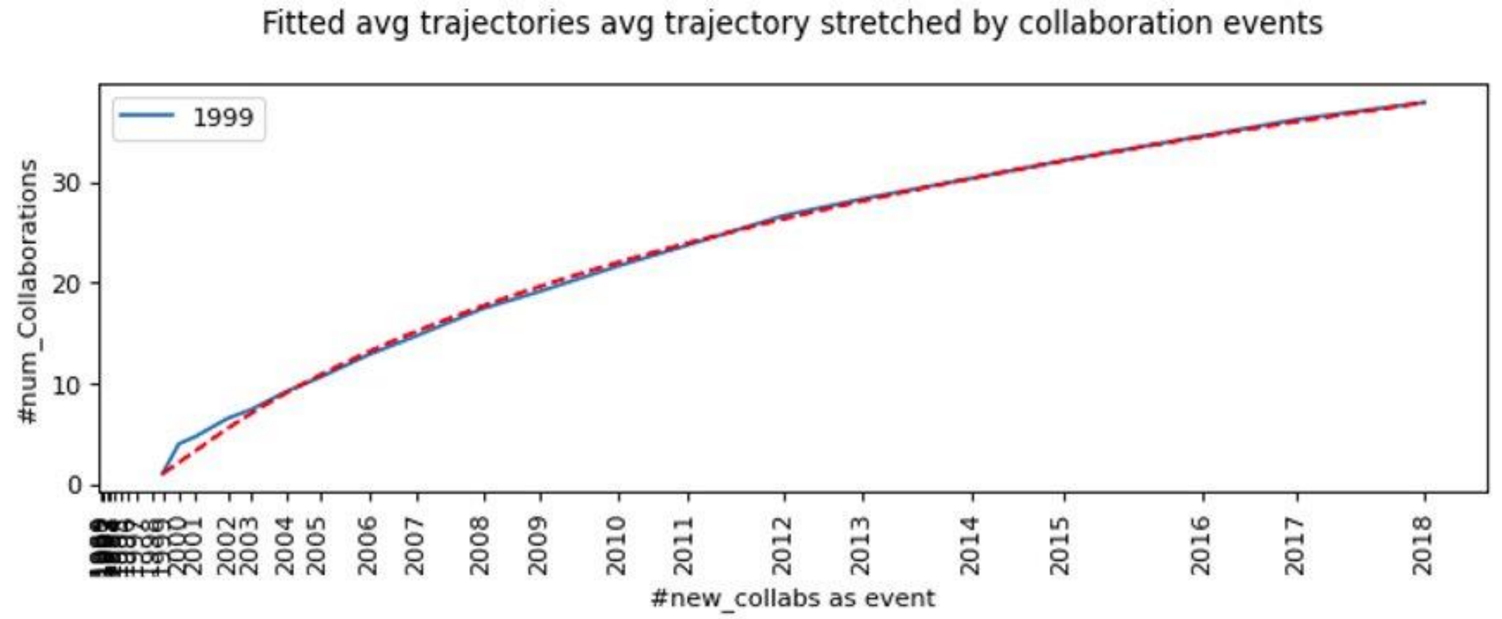
- Average trajectories are fitted one by one using the **logarithmic function** representing the theoretical vertex trajectory.

$$d_v(t) = (\alpha \ln(t/t_v) + 1)^{\beta}$$

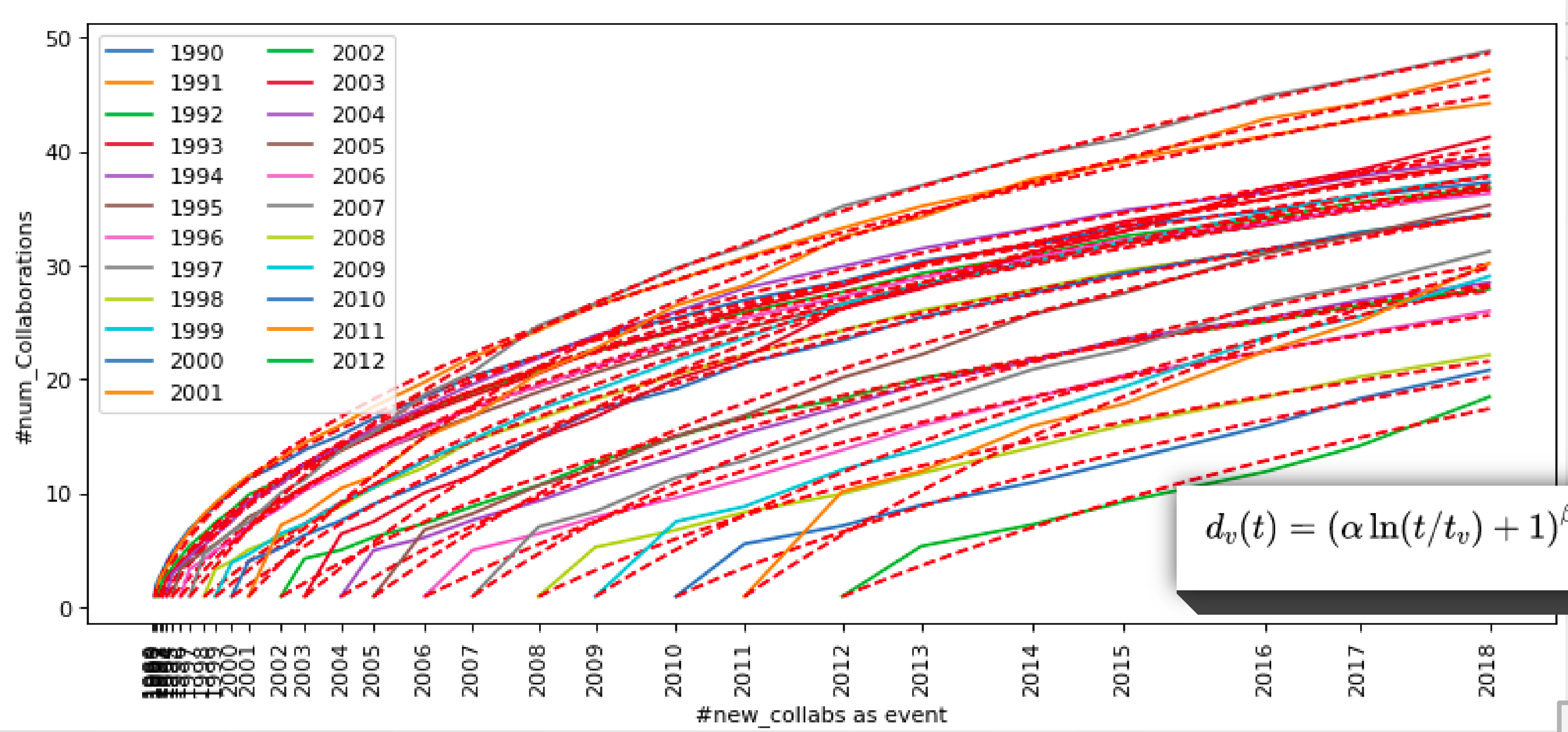
FITTING TRAJECTORIES

- Average trajectories are fitted one by one using the **logarithmic function** representing the theoretical vertex trajectory.
- Obtained a couple of individual parameter **α, β** for each.

$$d_v(t) = (\alpha \ln(t/t_v) + 1)^\beta$$



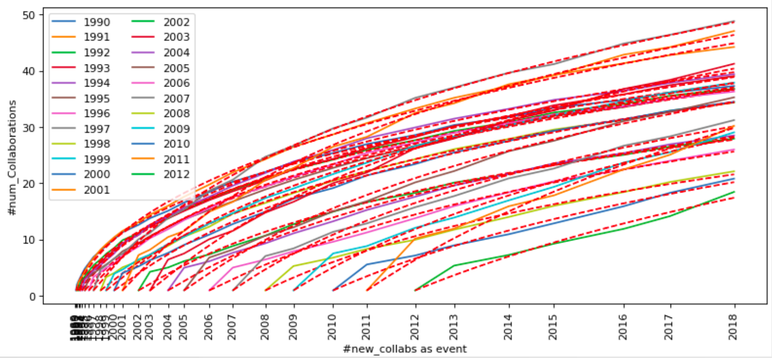
FITTING TRAJECTORIES



$$d_v(t) = (\alpha \ln(t/t_v) + 1)^\beta$$

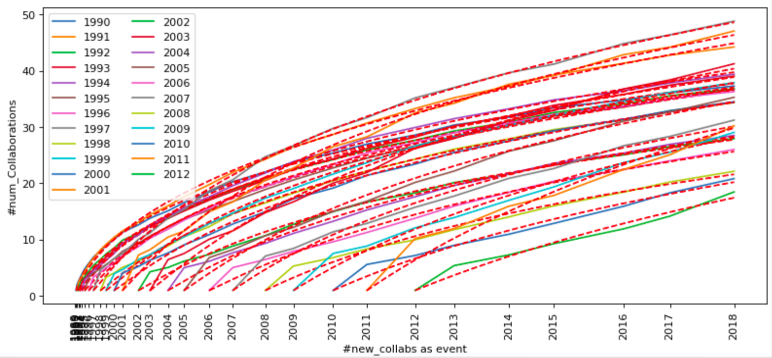
FITTING TRAJECTORIES

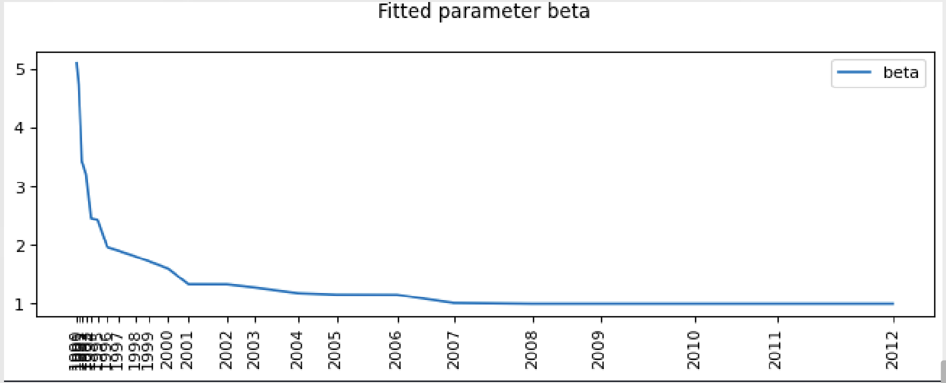
$$d_v(t) = (\alpha^* \ln(t/t_v) + 1)^{\beta^*}$$



FITTING TRAJECTORIES

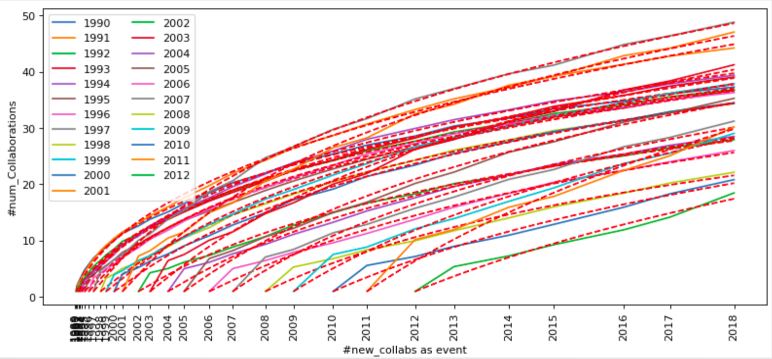
$$d_v(t) = (\alpha^* \ln(t/t_v) + 1)^{\beta^*}$$

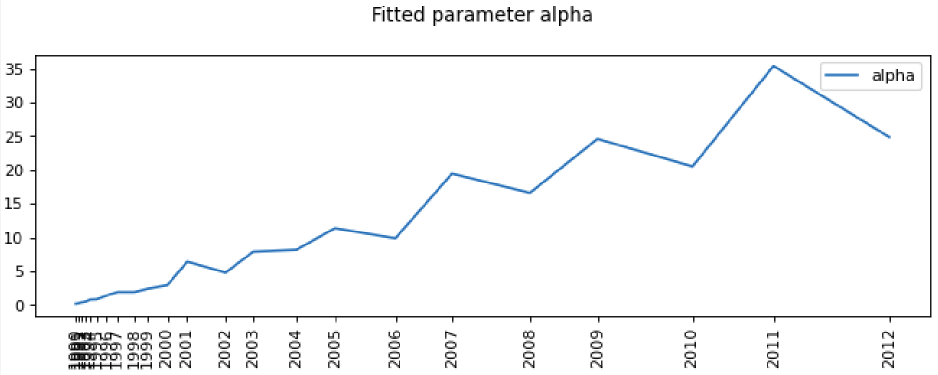


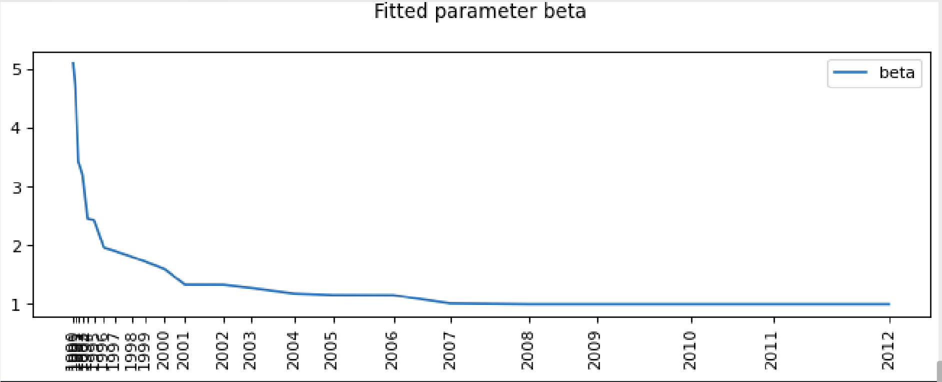


FITTING TRAJECTORIES

$$d_v(t) = (\alpha^* \ln(t/t_v) + 1)^{\beta^*}$$







GENERAL FITTING

- Find $d_v(t)=(\alpha^* \ln(t/t_v) + 1)^{\beta^*}$ | α^* **and** β^* fit all curves minimizing the given error.
- Four different kinds of errors:

GENERAL FITTING

- Find **$d_v(t)=(\alpha^* \ln(t/t_v) + 1)^{\beta^*}$ | α^* and β^*** fit all curves minimizing the given error.

- Four different kinds of errors:

A) $min_{\alpha^*,\beta^*}\left(\sum_{t_v}\sum_{t\geq t_v}|d_{t_v}(t)^* - r_{t_v}(t)|^2\right)$

B) $min_{\alpha^*,\beta^*}\left(\sum_{t_v}\max_{t\geq t_v}|d_{t_v}(t)^* - r_{t_v}(t)|^2\right)$

C) $min_{\alpha^*,\beta^*}\left(\sum_{t_v}\sum_{t\geq t_v}|d_{t_v}(t)^* - d_{t_v}(t)|^2\right)$

D) $min_{\alpha^*,\beta^*}\left(\sum_{t_v}\max_{t\geq t_v}|d_{t_v}(t)^* - d_{t_v}(t)|^2\right)$

GENERAL FITTING

- Find $d_v(t)=(\alpha^* \ln(t/t_v) + 1)^{\beta^*}$ | α^* **and** β^* fit all curves minimizing the given error.

- Four different kind of errors:

A) $min_{\alpha^*,\beta^*}\left(\sum_{t_v}\sum_{t\geq t_v}|d_{t_v}(t)^* - r_{t_v}(t)|^2\right)$

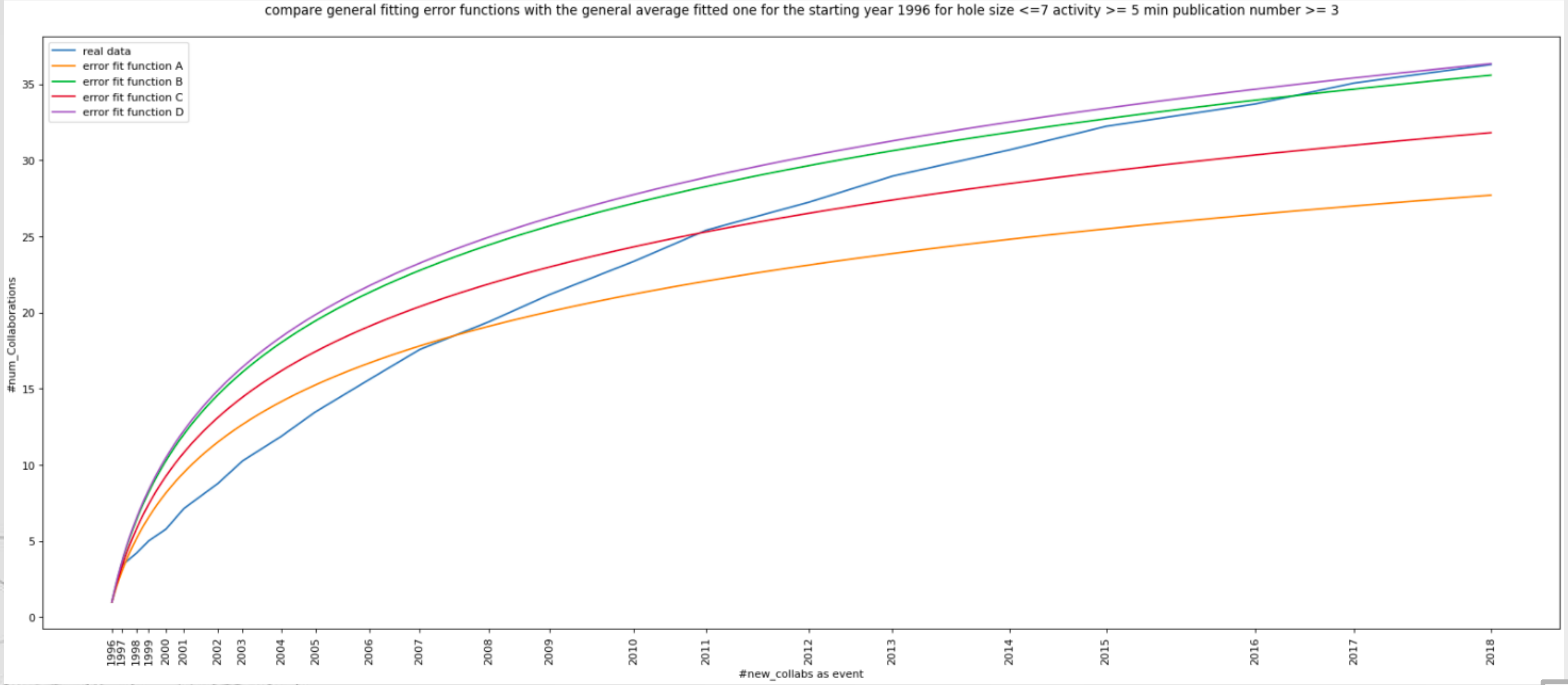
B) $min_{\alpha^*,\beta^*}\left(\sum_{t_v}\max_{t\geq t_v}|d_{t_v}(t)^* - r_{t_v}(t)|^2\right)$

C) $min_{\alpha^*,\beta^*}\left(\sum_{t_v}\sum_{t\geq t_v}|d_{t_v}(t)^* - d_{t_v}(t)|^2\right)$

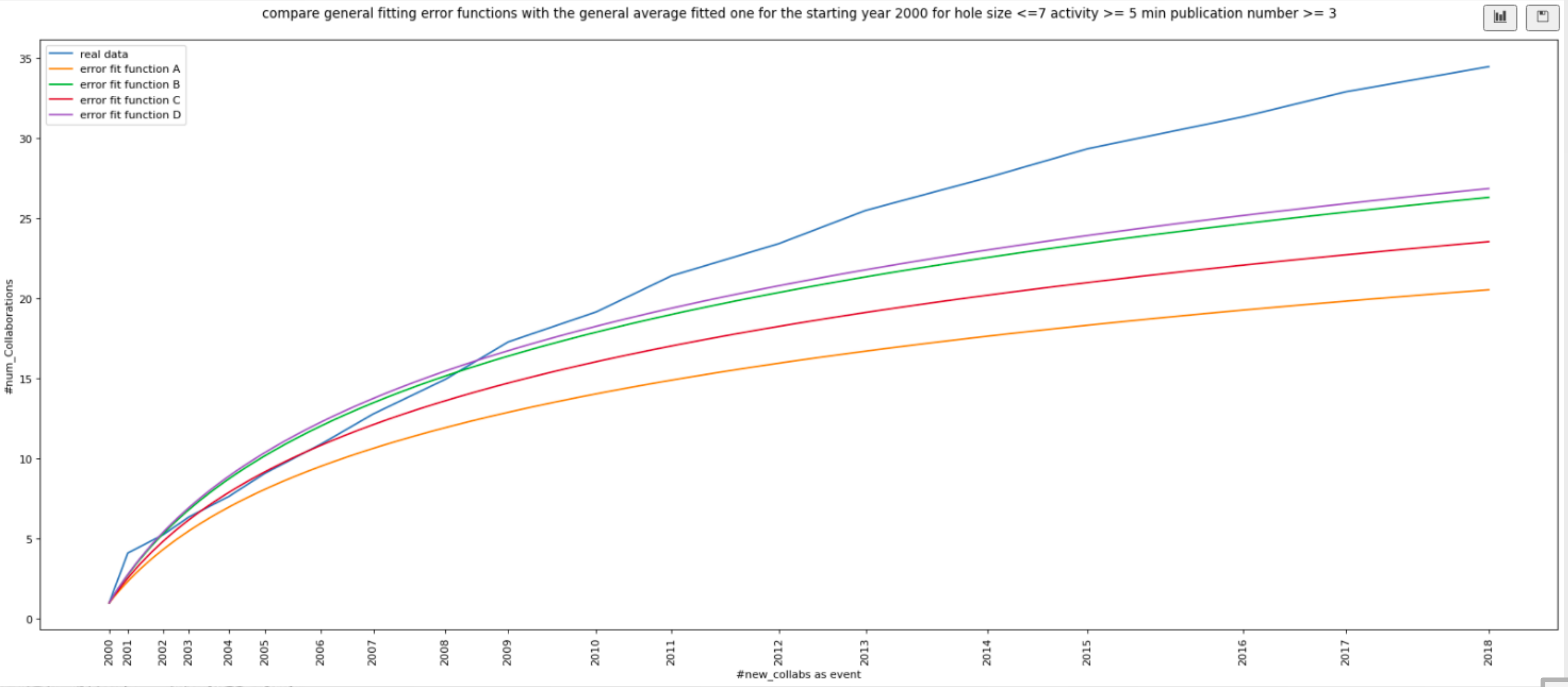
D) $min_{\alpha^*,\beta^*}\left(\sum_{t_v}\max_{t\geq t_v}|d_{t_v}(t)^* - d_{t_v}(t)|^2\right)$

	ERROR	alpha	beta
err_f_A	27755.784282	6.929222	1
err_f_B	4301.973888	8.973418	1
err_f_C	444951.754961	7.993638	1
err_f_D	4714.507436	9.169267	1

GENERAL FITTING (1996)



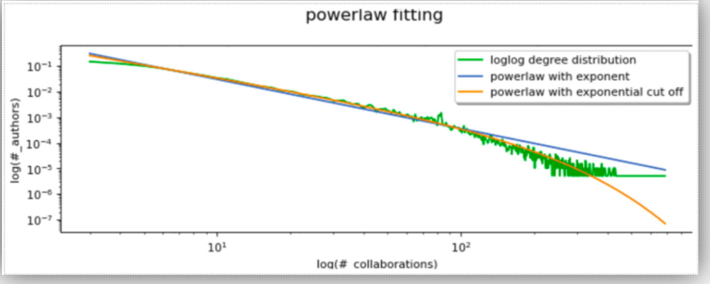
GENERAL FITTING (2000)



FINAL CONTRADDICTION

Real Data

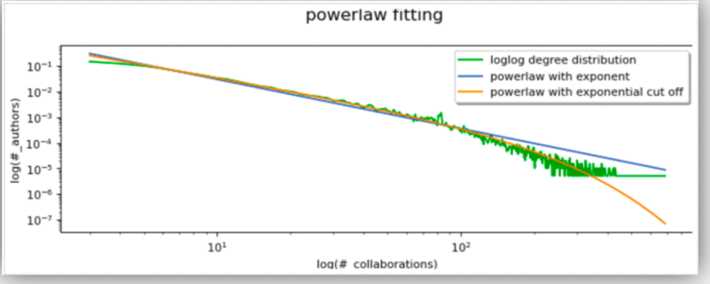
M_k



FINAL CONTRADDICTION

Real Data

M_k



Theoretical model

- $0 \leq \gamma < 1$
- $f(d_v(t)) = dv(t)^{\gamma}$

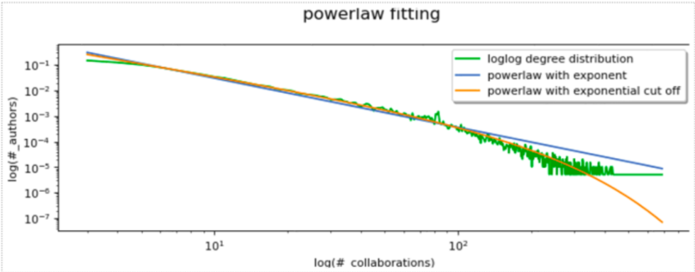
M_k

$$M_k \sim \alpha \cdot k^{-\gamma} \cdot \exp\left\{-\frac{\alpha}{1-\gamma} k^{1-\gamma}\right\}$$

FINAL CONTRADDICTION

Real Data

M_k



$d_v(t)$

$$d_v(t) = (\alpha \ln(t/t_v) + 1)^{\beta}$$

Theoretical model

- $0 \leq \gamma < 1$
- $f(d_v(t)) = dv(t)^{\gamma}$

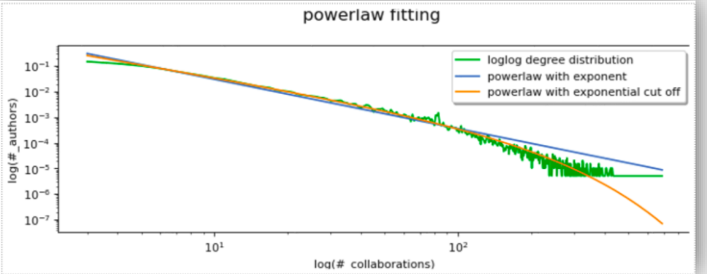
M_k

$$M_k \sim \alpha \cdot k^{-\gamma} \cdot \exp\left\{-\frac{\alpha}{1-\gamma} k^{1-\gamma}\right\}$$

FINAL CONTRADDICTION

Real Data

M_k



$d_v(t)$

$$d_v(t) = (\alpha \ln(t/t_v) + 1)^{\beta}$$

=> • β seems to be 1

Theoretical model

- $0 \leq \gamma < 1$
- $f(d_v(t)) = dv(t)^{\gamma}$

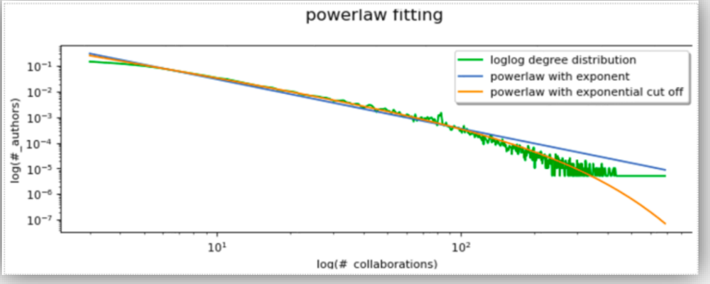
M_k

$$M_k \sim \alpha \cdot k^{-\gamma} \cdot \exp\left\{-\frac{\alpha}{1-\gamma} k^{1-\gamma}\right\}$$

FINAL CONTRADDICTION

Real Data

M_k



$d_v(t)$

$$d_v(t) = (\alpha \ln(t/t_v) + 1)^{\beta}$$

=> • β seems to be 1

Theoretical model

- $0 \leq \gamma < 1$
- $f(d_v(t)) = dv(t)^{\gamma}$

M_k

$$M_k \sim \alpha \cdot k^{-\gamma} \cdot \exp\left\{-\frac{\alpha}{1-\gamma} k^{1-\gamma}\right\}$$

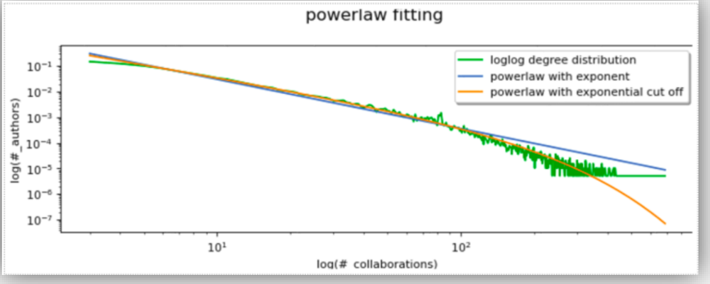
$d_v(t)$

$$d_v(t) = \left(\frac{1-\gamma}{\alpha}\ln(t/t_v) + 1\right)^{1/(1-\gamma)}$$

FINAL CONTRADDICTION

Real Data

M_k



$d_v(t)$

$$d_v(t) = (\alpha \ln(t/t_v) + 1)^{\beta}$$

- **β seems to be 1**

Theoretical model

- **$0 \leq \gamma < 1$**
- **$f(d_v(t)) = dv(t)^{\gamma}$**

M_k

$$M_k \sim \alpha \cdot k^{-\gamma} \cdot \exp\left\{-\frac{\alpha}{1-\gamma} k^{1-\gamma}\right\}$$

$d_v(t)$

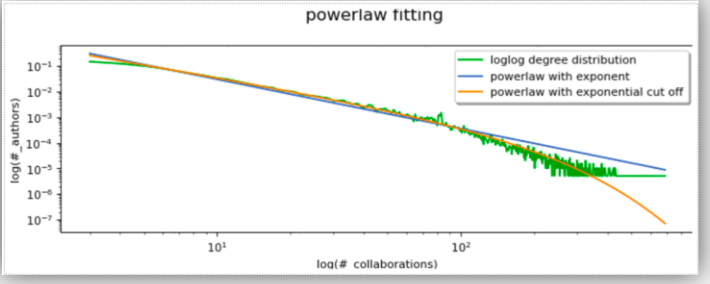
$$d_v(t) = \left(\frac{1-\gamma}{\alpha} \ln(t/t_v) + 1\right)^{1/(1-\gamma)}$$

=> $$\beta = \frac{1}{1-\gamma}$$

FINAL CONTRADDICTION

Real Data

M_k



$d_v(t)$

$$d_v(t) = (\alpha \ln(t/t_v) + 1)^{\beta}$$

- β seems to be 1

=> - $\gamma = 0$ and $f(d_v(t)) = 1$

Theoretical model

- $0 \leq \gamma < 1$
- $f(d_v(t)) = dv(t)^{\gamma}$

M_k

$$M_k \sim \alpha \cdot k^{-\gamma} \cdot \exp\left\{-\frac{\alpha}{1-\gamma} k^{1-\gamma}\right\}$$

$d_v(t)$

$$d_v(t) = \left(\frac{1-\gamma}{\alpha} \ln(t/t_v) + 1\right)^{1/(1-\gamma)}$$

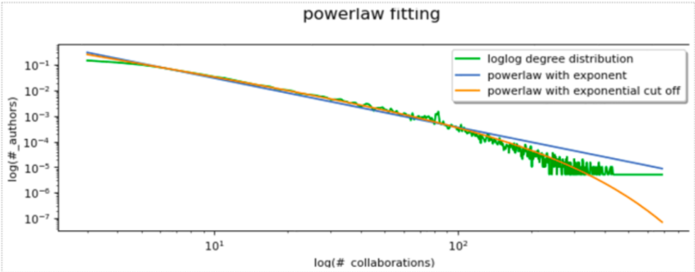
$$\beta = \frac{1}{1-\gamma}$$

=>

FINAL CONTRADDICTION

Real Data

M_k



$d_v(t)$ $d_v(t) = (\alpha \ln(t/t_v) + 1)^\beta$

- β seems to be 1
- γ = 0 and $f(d_v(t)) = 1$

Theoretical model

- $0 \leq \gamma < 1$
- $f(d_v(t)) = dv(t)^\gamma$

M_k

$$M_k \sim \alpha \cdot k^{-\gamma} \cdot \exp\left\{-\frac{\alpha}{1-\gamma} k^{1-\gamma}\right\}$$

$d_v(t)$

$$d_v(t) = \left(\frac{1-\gamma}{\alpha}\ln(t/t_v) + 1\right)^{1/(1-\gamma)}$$

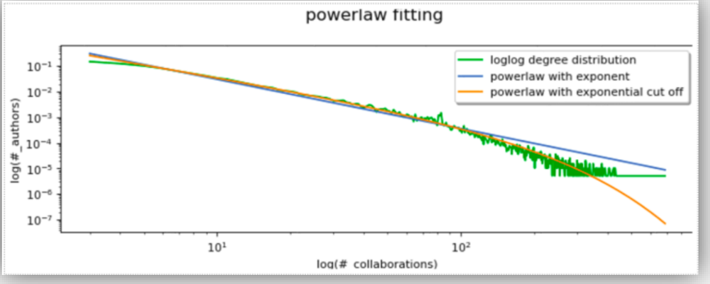
$$\beta = \frac{1}{1-\gamma}$$

=> $$\Pr[v \text{ is chosen at time step t}] = \frac{f\left(\mathrm{d}_t(v)\right)}{\sum_{w \in V_t} f\left(\mathrm{d}_t(w)\right)}$$

FINAL CONTRADDICTION

Real Data

M_k



$d_v(t)$

$$d_v(t) = (\alpha \ln(t/t_v) + 1)^{\beta}$$

- **β seems to be 1**
- **$\gamma = 0$ and $f(d_v(t)) = 1$**

Theoretical model

- **$0 \le \gamma < 1$**
- **$f(d_v(t)) = dv(t)^{\gamma}$**

M_k

$$M_k \sim \alpha \cdot k^{-\gamma} \cdot \exp\left\{-\frac{\alpha}{1-\gamma} k^{1-\gamma}\right\}$$

$d_v(t)$

$$d_v(t) = \left(\frac{1-\gamma}{\alpha} \ln(t/t_v) + 1\right)^{1/(1-\gamma)}$$

$$\beta = \frac{1}{1-\gamma}$$

$$\Pr[v \text{ is chosen at time step t}] = \frac{1}{n}$$ **<=**

06

NEXT STEPS

NEXT STEPS

The underlying theoretical model may be different from the one we assumed at the beginning:

Next steps:

1. **Generating a network** on a given theoretichal model **($f(d_v(t)=1)$** to fit our data;
2. Generating slightly different networks to check **which represent better** our case;
3. Defining a **new theorethical model** describing better the discovered behavior [11];
4. The analysis of data from **other research field**;

References:

1. Derek J. de Solla P. **Networks of scientific papers**, 1965, doi:10.1126/science.149.3683.510.

2. Michalis F. Petros F., and Christos F. **On power-law relationships of the Internet topolog**y, 1999, doi:10.1145/316188.316229.

3. Bollobás B. and Riordan O**. Handbook of Graphs and Networks: From the Genome to the Internet,** 2003. Pages 1–34.

4. *Broido A. D.and A. Clauset,* ***"Scale-free networks are rare",*** 2019, doi: 10.1038/s41467-019-08746-5.

5. *Newman M. E. J.* ***Coauthorship networks and patterns of scientific collaboration,*** 2004, doi:10.1073/pnas.0307545100.

6. *Newman M. E. J.* ***Clustering and preferential attachment in growing networks,*** 2001, doi:10.1103/PhysRevE.64.025102.

7. Barabasi A. and Albert R. **Emergence of Scaling in Random Networks.** 1999, Pages 509–12, doi: 10.1126/science.286.5439.509.5

8. *Giroire F., Nisse N., Sulkowska M.,* ***Study of a degree distribution and a vertex trajectory in the Chung-Lu model with a generalized attachment function,*** *2022*

9. *Giroire F., Nisse N.,*Thibaud T, *Sulkowska M.* . **Preferential attachment hypergraph with high modularity**, 2021, arXiv:2103.01751.

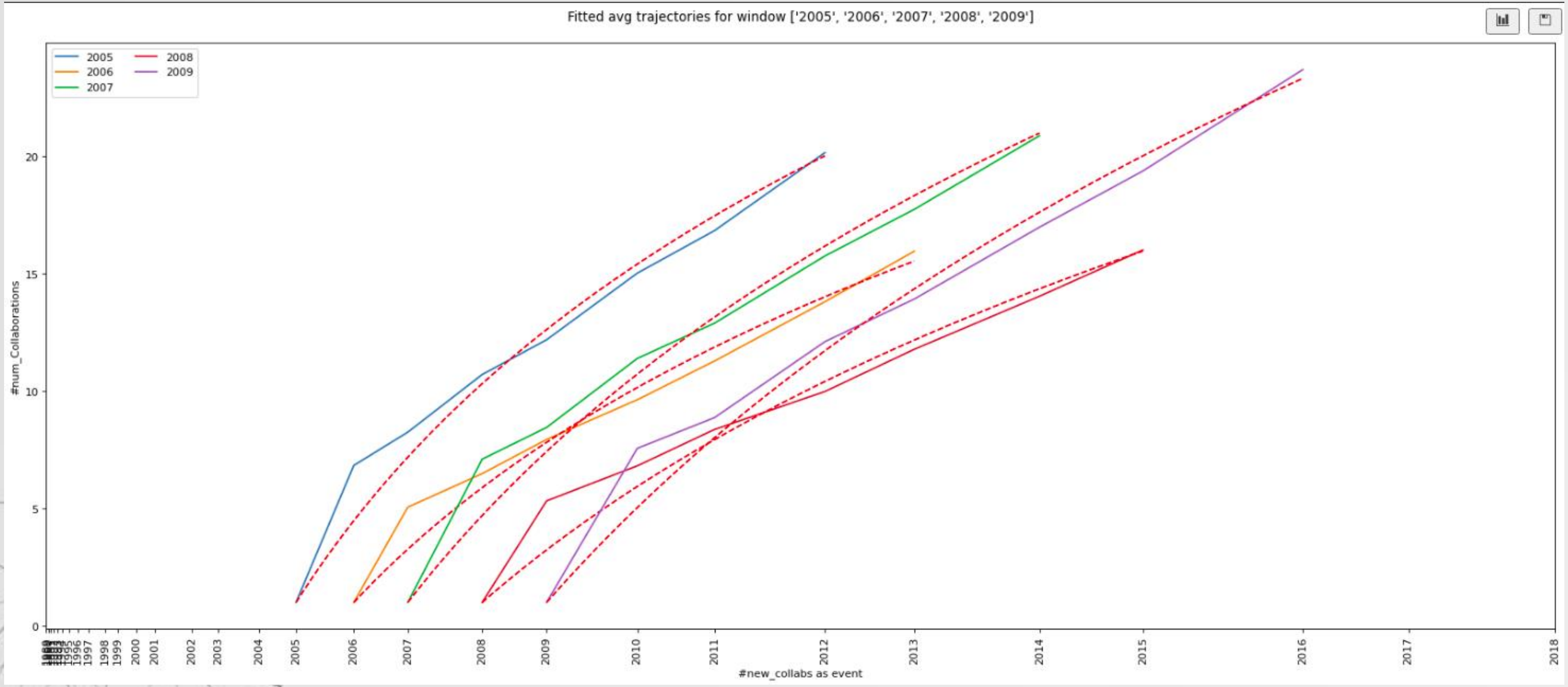
10. Chung F. and Lu L. **Complex Graphs and Networks**, 2006.

11. Cooper C. , Alan F. **A General Model of Web Graphs, 2001**


The End

WINDOW-SECTION FITTING

The same procedure is applied to **five year windows** and considering only the first **eight year section** of each trajectory.



WINDOW-SECTION FITTING

The general fitting this time is performed of the A and C errors type defined previously obtain the following parameters.

window	alpha	beta	error A
[1989, 1990, 1991, 1992, 1993]	0.642520	1.000000	315.834775
[1990, 1991, 1992, 1993, 1994]	2.479811	1.039867	45.397347
[1991, 1992, 1993, 1994, 1995]	3.227549	1.000000	38.597747
[1992, 1993, 1994, 1995, 1996]	3.733596	1.000000	48.907064
[1993, 1994, 1995, 1996, 1997]	4.625937	1.000000	113.231100
[1994, 1995, 1996, 1997, 1998]	5.258093	1.000000	88.733436
[1995, 1996, 1997, 1998, 1999]	6.043890	1.000000	86.015274
[1996, 1997, 1998, 1999, 2000]	6.907860	1.000000	59.308019
[1997, 1998, 1999, 2000, 2001]	8.611239	1.000000	221.401225
[1998, 1999, 2000, 2001, 2002]	8.964712	1.000000	213.152739
[1999, 2000, 2001, 2002, 2003]	10.635242	1.000000	272.652060
[2000, 2001, 2002, 2003, 2004]	11.634065	1.000000	210.052291
[2001, 2002, 2003, 2004, 2005]	13.394332	1.000000	177.619443
[2002, 2003, 2004, 2005, 2006]	13.390073	1.000000	174.522331
[2003, 2004, 2005, 2006, 2007]	15.449380	1.000000	144.535014
[2004, 2005, 2006, 2007, 2008]	15.642647	1.000000	144.264338
[2005, 2006, 2007, 2008, 2009]	17.902454	1.000000	210.980700
[2006, 2007, 2008, 2009, 2010]	18.600528	1.000000	208.848946
[2007, 2008, 2009, 2010, 2011]	21.851542	1.000000	414.262737

window	alpha	beta	error C
[1989, 1990, 1991, 1992, 1993]	1.680922	1.000000	116469.182190
[1990, 1991, 1992, 1993, 1994]	2.795480	1.117923	4744.905551
[1991, 1992, 1993, 1994, 1995]	3.054005	1.100979	3556.120018
[1992, 1993, 1994, 1995, 1996]	4.164666	1.000000	5641.857099
[1993, 1994, 1995, 1996, 1997]	4.857185	1.000000	16296.886017
[1994, 1995, 1996, 1997, 1998]	5.392140	1.000000	12396.764940
[1995, 1996, 1997, 1998, 1999]	5.968222	1.000000	14108.820976
[1996, 1997, 1998, 1999, 2000]	6.825733	1.000000	7909.599641
[1997, 1998, 1999, 2000, 2001]	8.262750	1.000000	26662.213671
[1998, 1999, 2000, 2001, 2002]	8.596449	1.000000	26030.916200
[1999, 2000, 2001, 2002, 2003]	10.192800	1.000000	29151.731420
[2000, 2001, 2002, 2003, 2004]	11.310162	1.000000	21234.077434
[2001, 2002, 2003, 2004, 2005]	13.104896	1.000000	13719.973595
[2002, 2003, 2004, 2005, 2006]	12.935749	1.000000	13522.175020
[2003, 2004, 2005, 2006, 2007]	15.101792	1.000000	6540.280182
[2004, 2005, 2006, 2007, 2008]	15.176833	1.000000	6653.254309
[2005, 2006, 2007, 2008, 2009]	1.961128	2.388138	22719.342323
[2006, 2007, 2008, 2009, 2010]	17.747342	1.000000	6585.673234
[2007, 2008, 2009, 2010, 2011]	20.972356	1.000000	9226.825981