The Role of Research Leaders on the Evolution of Scientific Communities

Bruno Leite Alves

Fabrício Benevenuto   Alberto H. F. Laender

Federal University of Minas Gerais
Belo Horizonte, Brazil

- My name is Bruno Leite Alves and I will talk about the role of research leaders on the evolution of scientific communities.

- This work was done with Fabrício Benevenuto and Alberto Laender.

- We are from Federal University of Minas Gerais, from Belo Horizonte, Brazil.

Society is Organized in Communities

- There are a lot of kinds of communities

- In a social networks, individuals influence and are influenced by others

- Communities have opinion leaders

A group of leaders or influential members are able to affect the dynamics of the entire community

- The society is organized in communities and there are a lot of kinds of communities, such as: fans of a sport, fans of a celebrity, friend's communities, and scientific communities and so on.

- We may name these communities as social networks, and inside of these social networks there are individuals who influence and are influenced by others.

- There are a lot of works which point that these communities have leaders and these leaders are able to affect the opinion of other people. In this way, groups of leaders or influential are able to affect the dynamics of the entire communities.

Our Goal

- Study the dynamics of **scientific communities**

- Identify leaderships within **scientific communities** (Community Core)

- Investigate the properties of a Community Core

- Here, we want to identify leaders inside of scientific communities.
- In this way, our goal is to study the dynamics of scientific communities and identify leaderships in these communities, here we named community core.
- We are also interested in the properties of the community core.

Outline

- Data Source about Scientific Communities

- Extracting the Community Core

- Properties of a Community Core

- So, the rest of the speech is organized as it follows.

- Next, I am going to describe the data we used to study scientific communities.

- Then, we describe how we extract the core of a community, and finally we present a series of analysis about the communities core.

Data Source

2.2 million publications

from **1.2** million authors

24 flagship conferences

of **major ACM SIGs**

We considered each **conference**

as a **scientific community**

- We used the DBLP data, a digital library with 2.2 million publications from 1.2 million authors.

- As the DBLP has a lot of information, for a better analysis, we selected 24 flagship conferences of major ACM SIGs and we considered each conference as a scientific community.

ACM SIGs Considered

SIGACT	SIGDOC	SIGMOD
SIGAPP	SIGGRAPH	SIGOPS
SIGARCH	SIGIR	SIGPLAN
SIGBED	SIGKDD	SIGSAC
SIGCHI	SIGMETRICS	SIGSAM
SIGCOMM	SIGMICRO	SIGSOFT
SIGCSE	SIGMM	SIGUCCS
SIGDA	SIGMOBILE	SIGWEB

- So, here it is possible to observe the 24 SIGs selected, all of our analysis are based in these conferences.

- We have selected conferences with 10 years or more.

Core Score

- Estimates a researcher's importance within a community.

- The core score of a researcher *r* into a community *c* in a period of time *t* is given by:

$$CoreScore_{r,c,t} = h_r \times \#publications_{r,c,t}$$

- In this context, which we would like to identify the most important members, we needed to quantify the importance of the researchers inside of communities.
- So we defined this metric: Core Score. The core score estimates a researcher's importance within a community and the core score of a researcher r into a community c in a specific time t is given by h-index of a researcher times the number of publications of a researcher into a community in a period of time.
- An important piece of this metric is the h-index, so we needed to obtain this value.

How to Estimate H-index

Only **30%**

of **DBLP** authors had a profile at

Google Scholar

- So, our first attempt (try) it was to use the Google Scholar, an important tool which is widely used among scientific communities to calculate the h-index.
- For it, we randomly selected 10 researchers for each conference, but only 30% of DBLP authors had a profile at Google Scholar.

How to Estimate H-index

Only **30%**

of **DBLP** authors had a profile at

Google Scholar

Alternative

SHINE
SIMPLE HINDEX ESTIMATION

www.shine.icomp.ufam.edu.br

- Then we used Shine data.

- Shine project is a tool which makes the estimation of h-index of conferences.

- They use the number of references of the papers to estimate the h-index, so we match this information with DBLP and we infer the h-index of the researchers.

Shine vs. Google Scholar
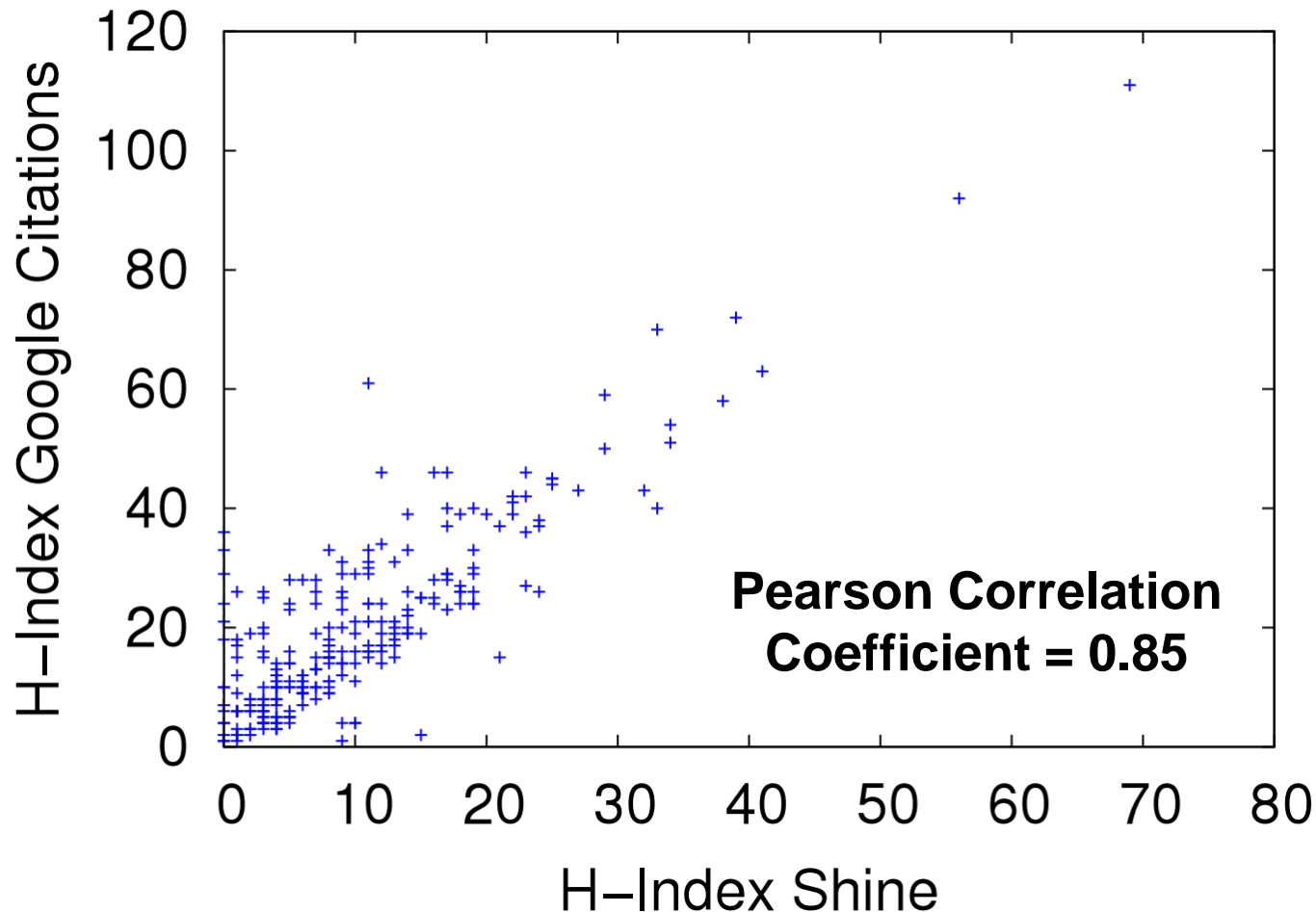
Shine offers a good estimation for H-Index

- To validate this estimative, we calculated the Pearson correlation coefficient.

- It is possible to note in the graphic and the value of Pearson Coefficient a strong positive correlation between the inferred value and Google Scholar value.

- So, Shine offers a good estimation for H-Index.

Resemblance and Angular Coefficient
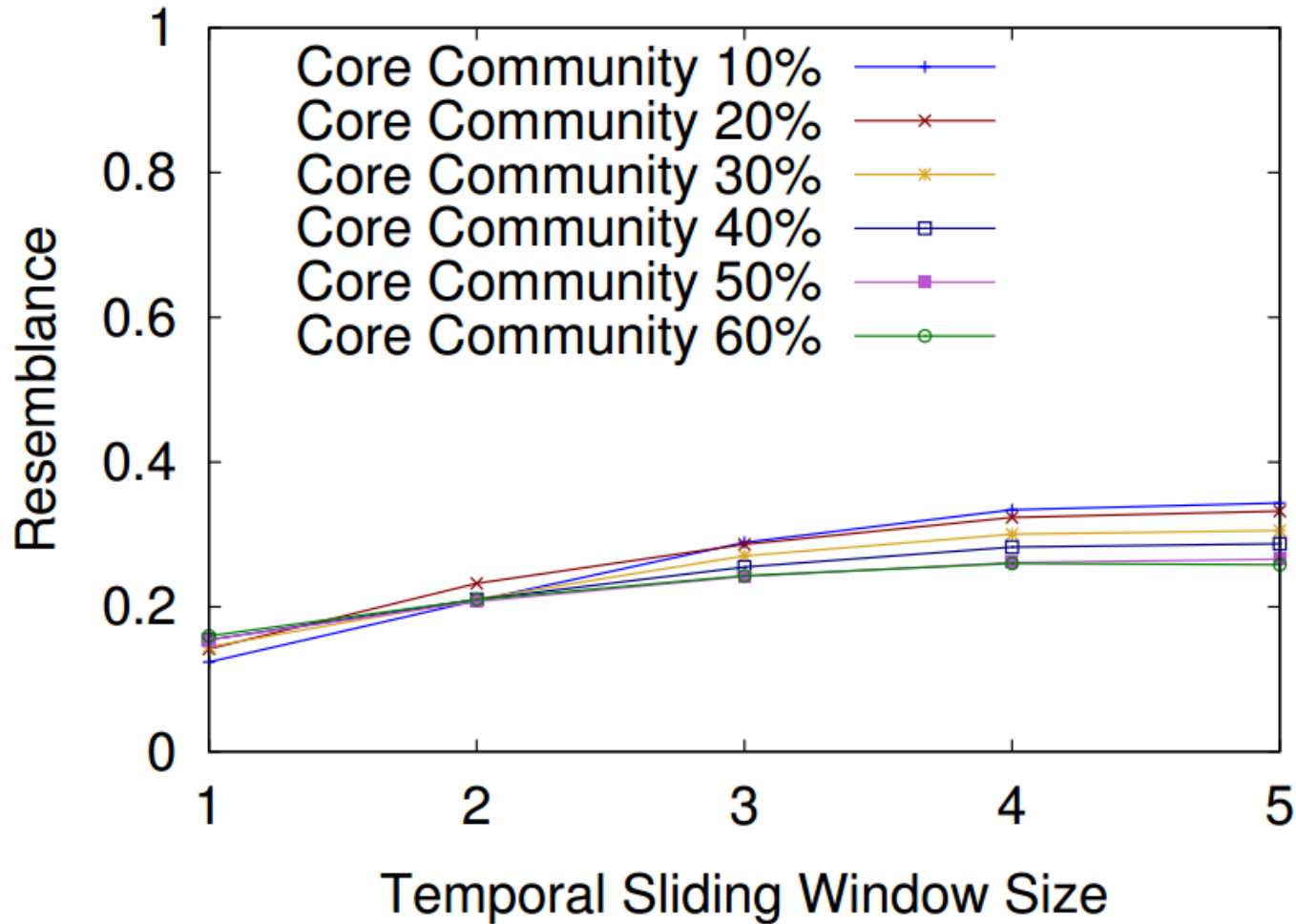
SIGMOD

- After it was done, we needed to define two important thresholds to extract the core community.

- The first one is the size of core community, once we have a ranking based in the core score.

- The second one is the temporal sliding window size to analyze the core community over the time.

- We used the resemblance and angular coefficient to help us. The resemblance is a measure which identifies the change in the network over the time and angular coefficient is responsible to identify the inclination of the lines.

Resemblance and Angular Coefficient
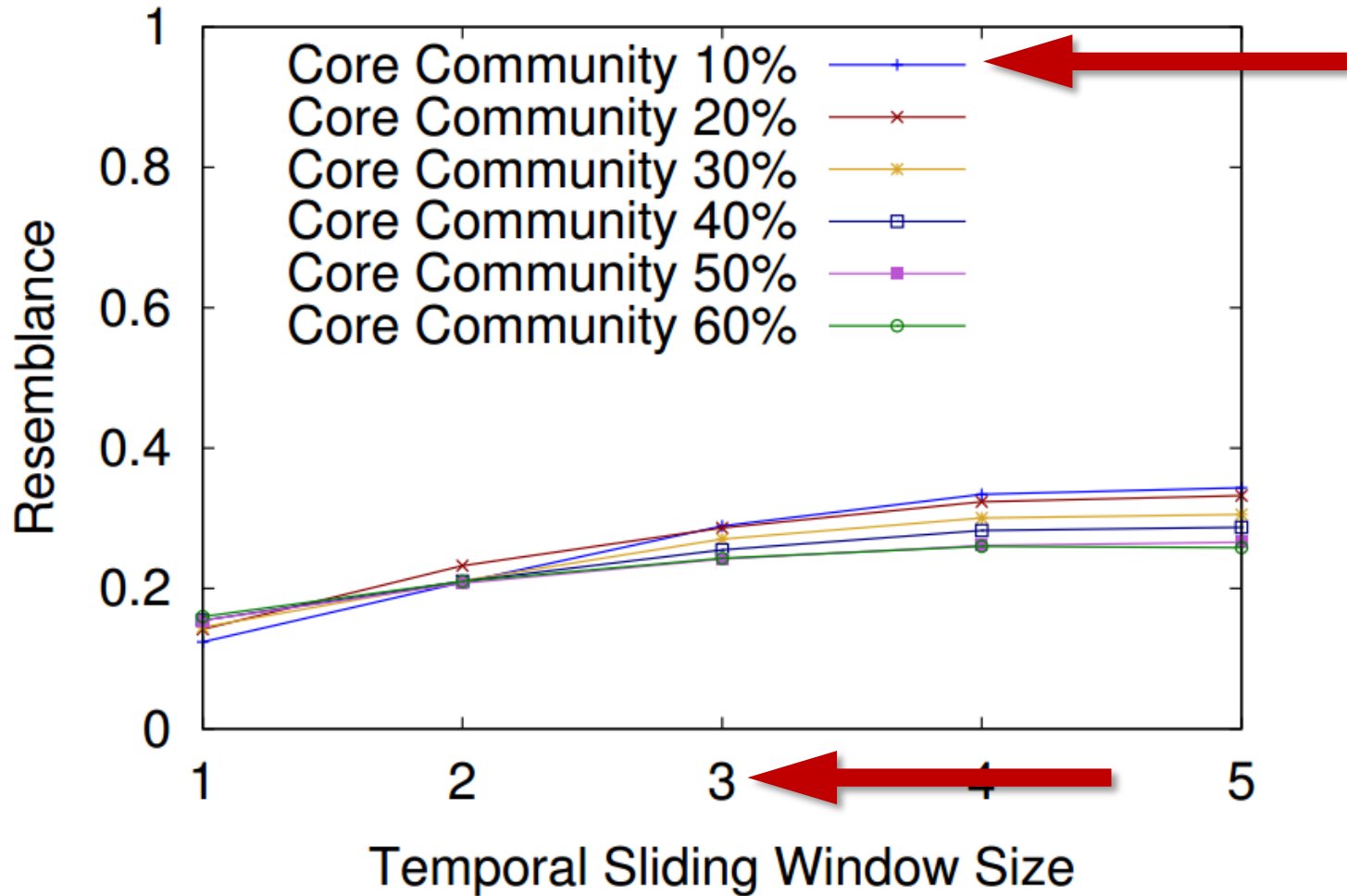
SIGMOD

- After the calculations we defined the core community size as 10% and the temporal sliding window size as 3 years. Here it is possible to observe the SIGMOD conference, but the other conferences follow the same behavior.

Luis von Ahn's Communities
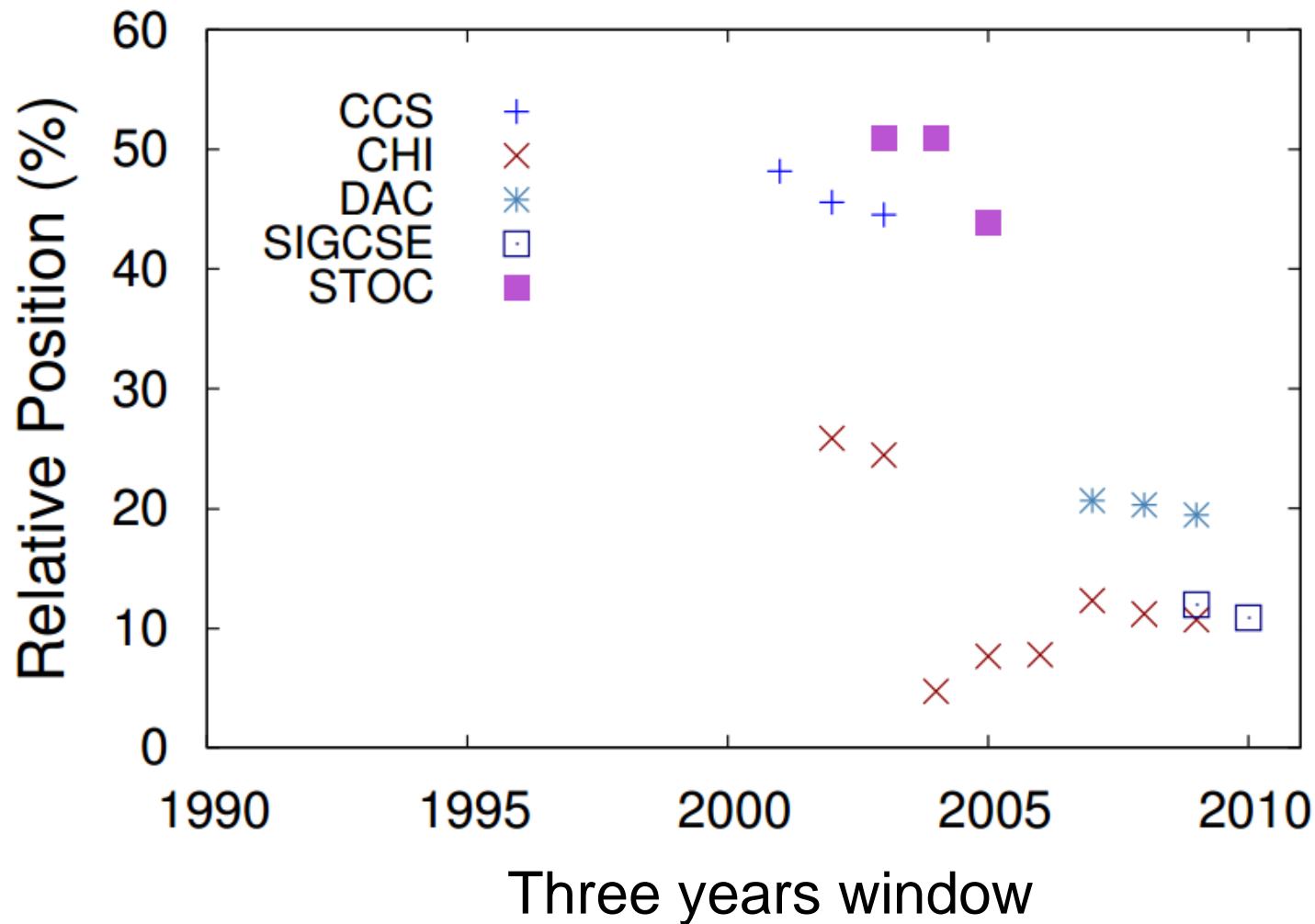
High core score in CHI and SIGCSE

- We would like to validate the ranking of the core score, so we selected two important researchers of scientific communities, the first one is Luis von Ahn, the keynote speaker of this conference this year. Here, we may note the core score of Luis von Ahn's communities.
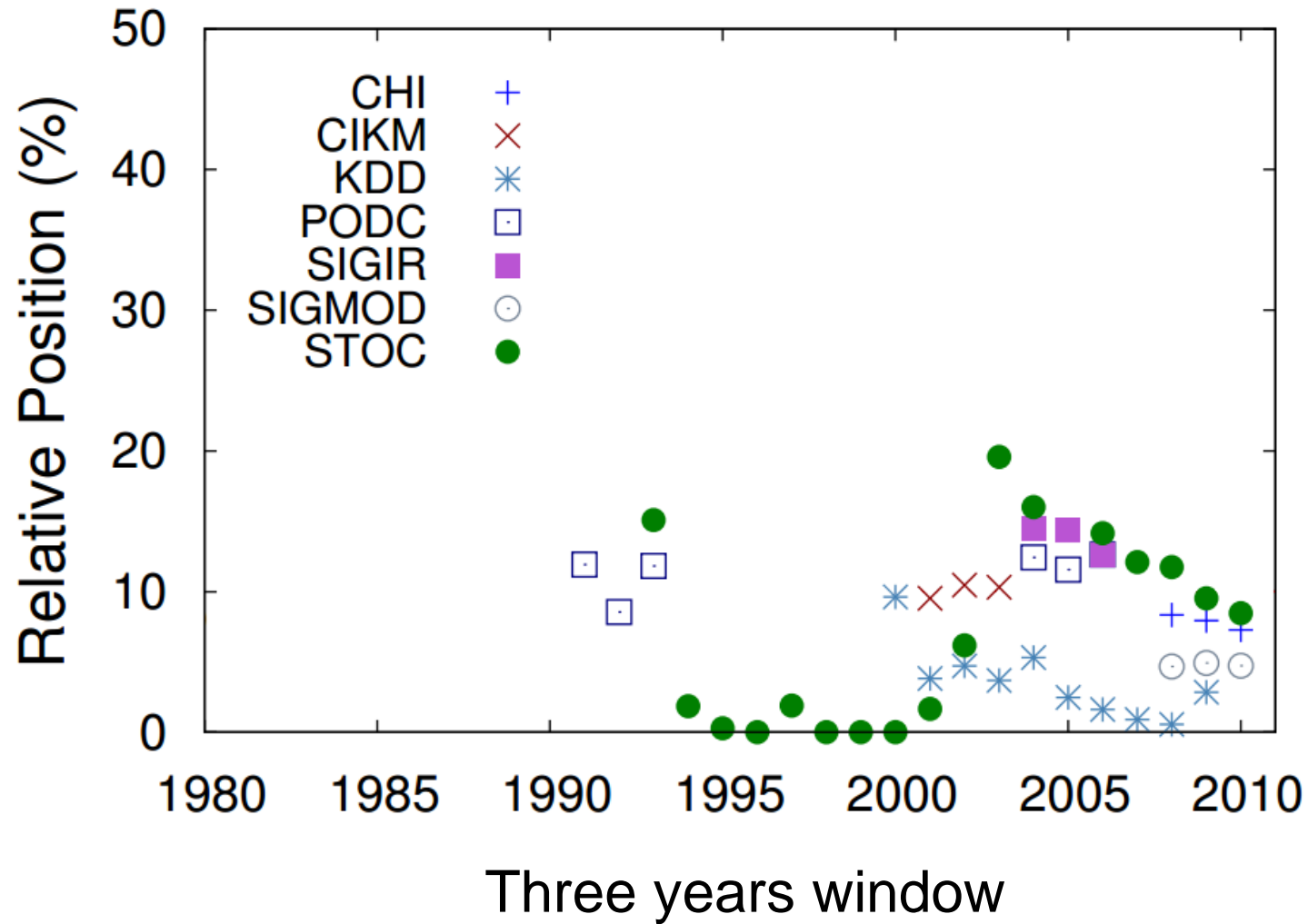
- Luis has a high core score in CHI and SIGCSE.

Luis von Ahn's Communities
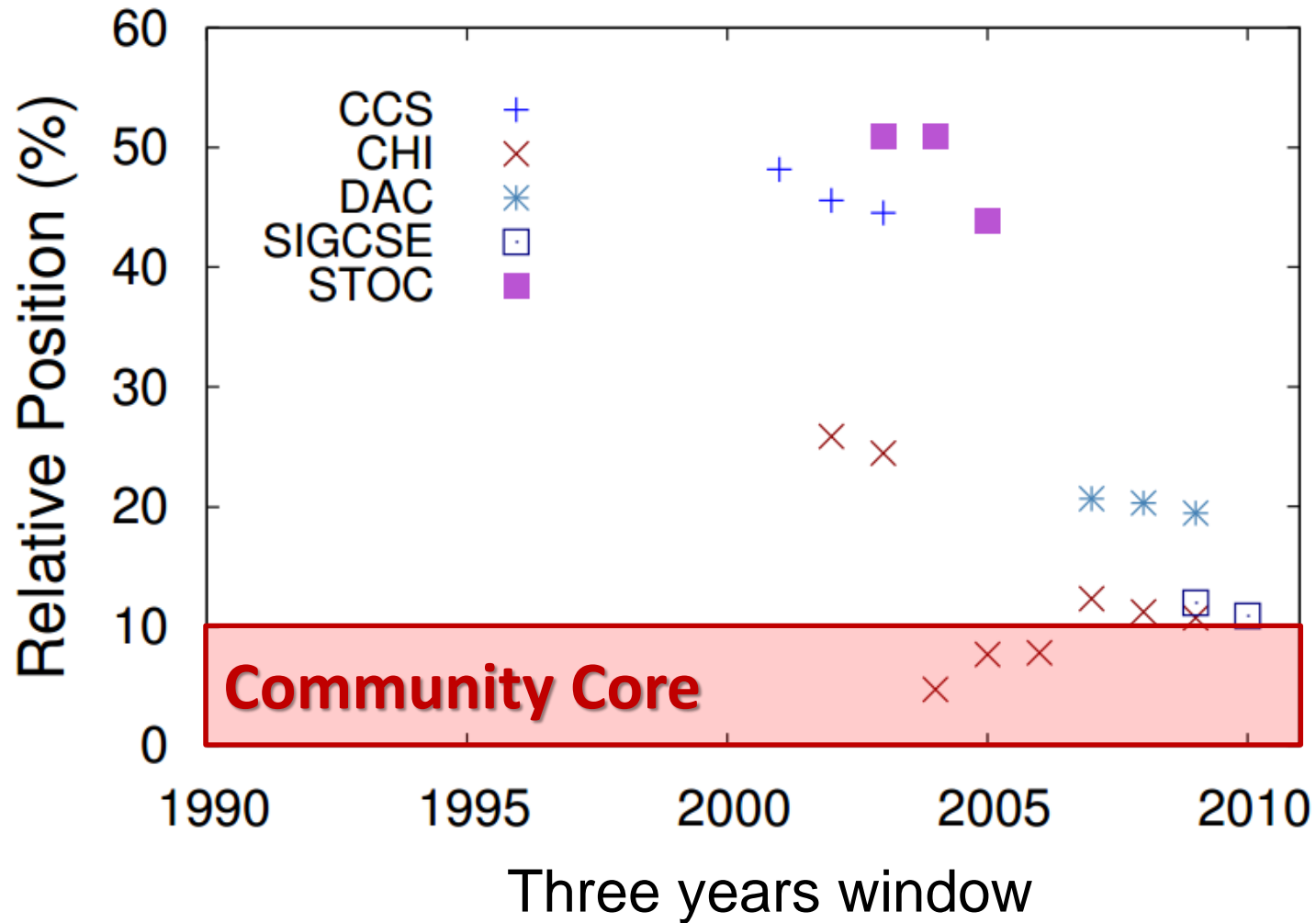
Member of the Community Core of CHI

- We may see here that Luis became part of community core of CHI and he was close to become part of the community core of conferences as SIGCSE and after DAC.
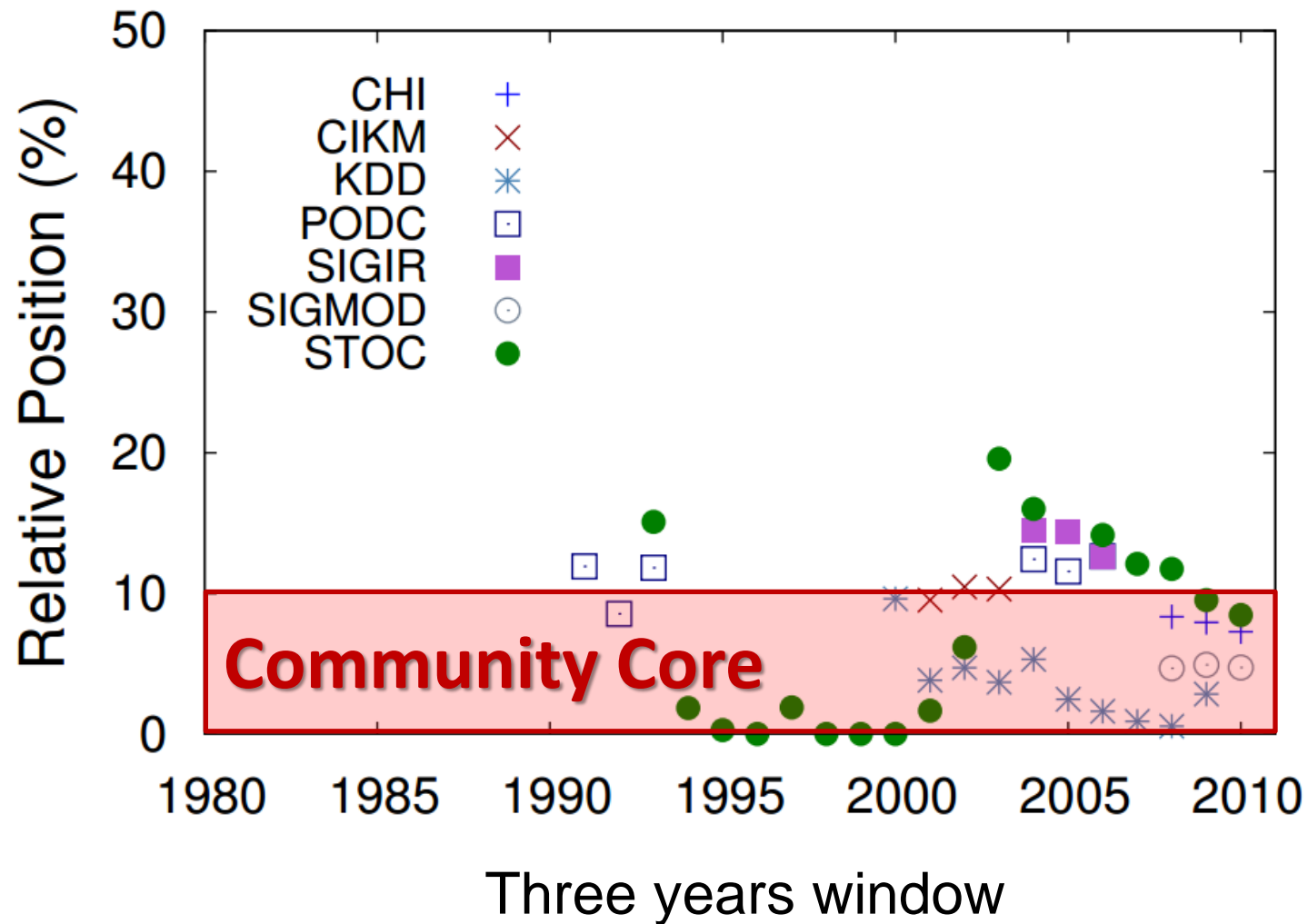
Jon Kleinberg's Communities

Three years window

High core score in several communities

- Other important researcher and keynote speaker is Jon Kleinberg (Klaiber). We may observe Jon Kleinberg as member of the community core of several conferences.
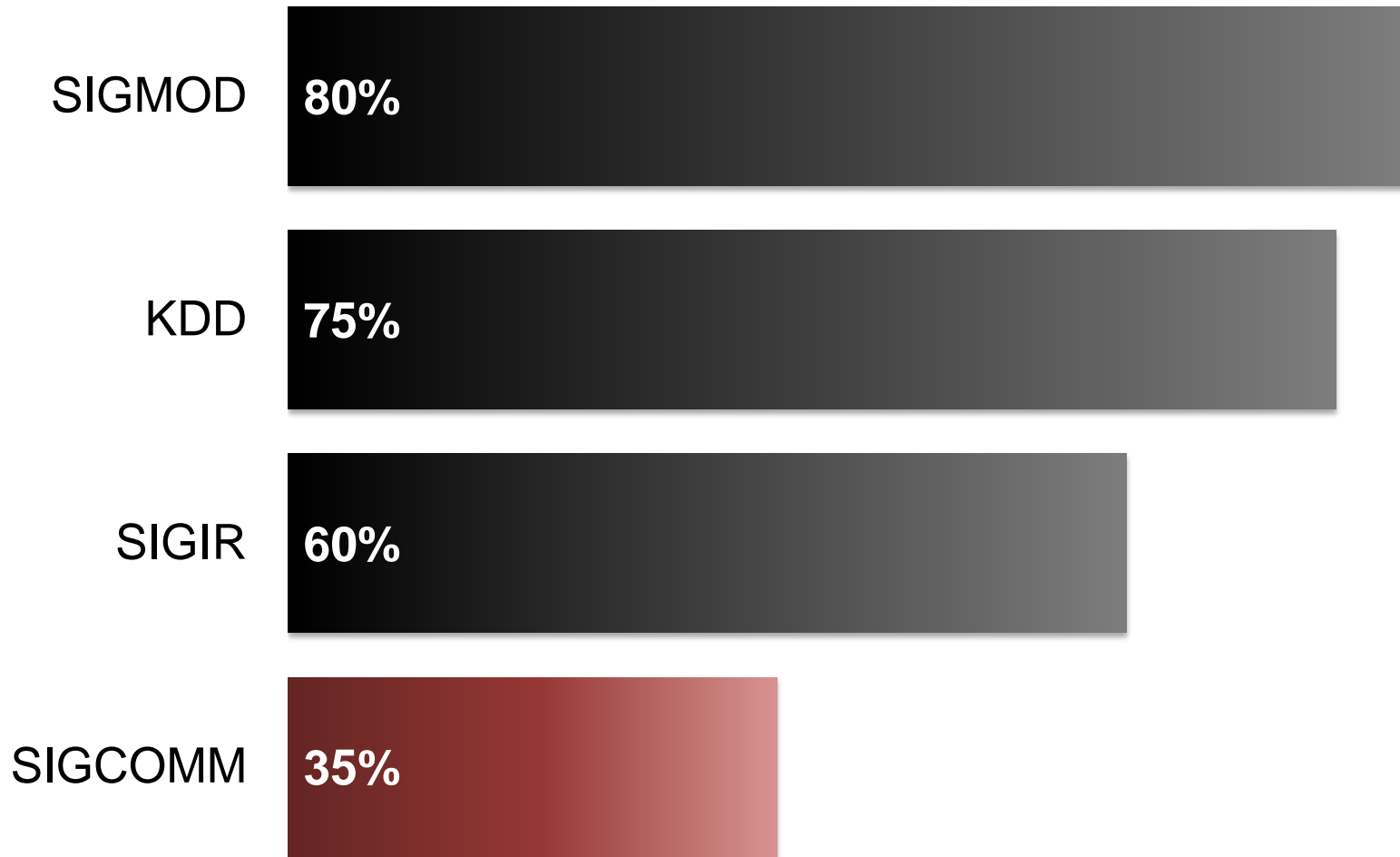
Jon Kleinberg's Communities

Three years window

Moving from STOC to KDD

- It is possible note Jon Kleinberg moving from STOC to KDD and other conferences.

Awarded Researchers

SIGMOD   **80%**

KDD   **75%**

SIGIR   **60%**

SIGCOMM   **35%**

Members of the core community awarded

- A second validate we proposed is to identify conference which was awarded as the best researcher and observe whether these researchers were members of the community core.

- It is possible to observe a high number of members of community core awarded in SIGMOD, KDD and SIGIR.

- In SIGCOMM case, several researchers appeared in SIGCOMM sponsors conference.

Evolution of the Scientific Communities
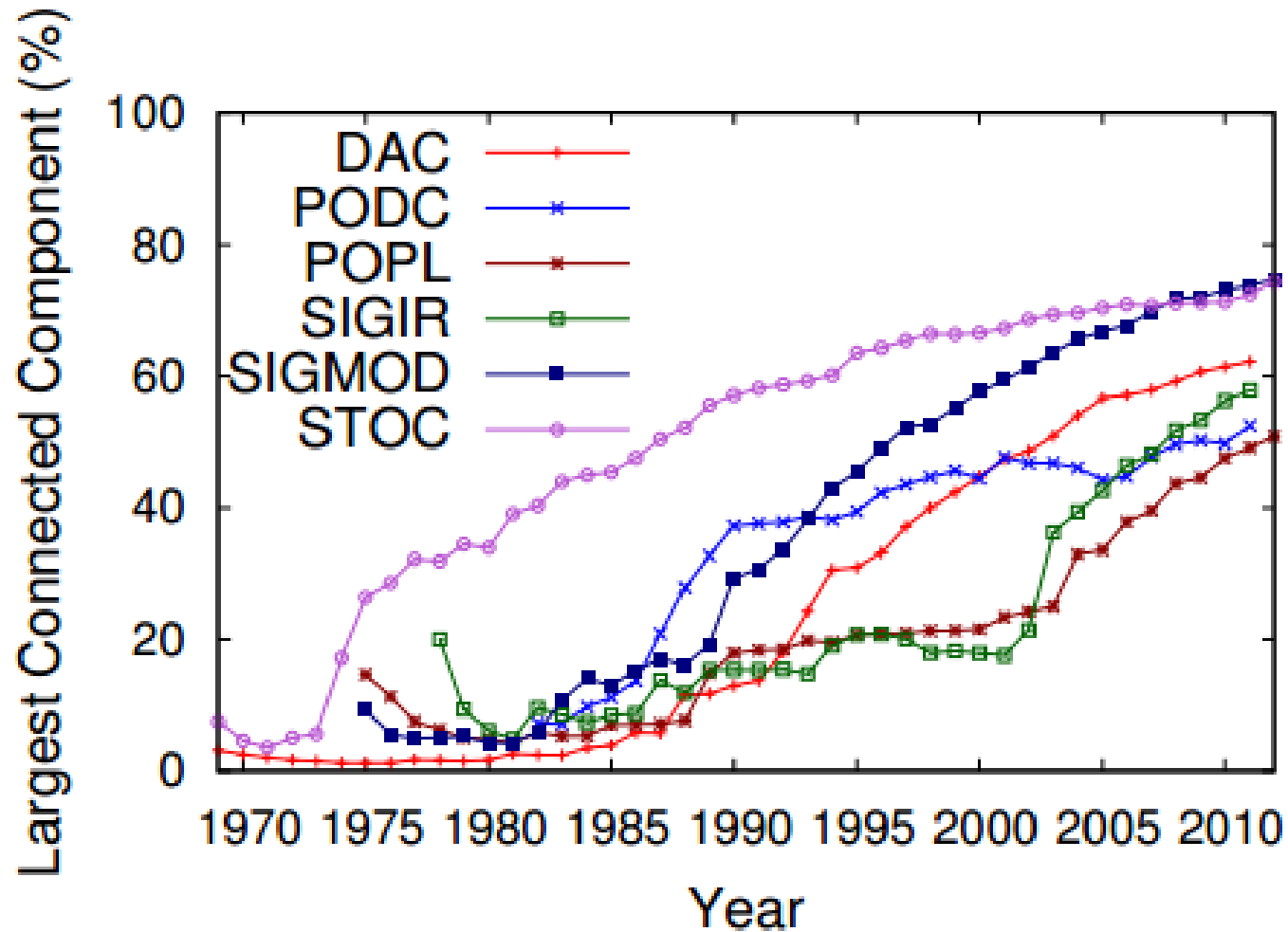
Year by year by accumulating nodes and edges

- Our analyses are performed under two perspectives.
- The first consists on analyzing the network evolution year by year by accumulating nodes and edges to a single final snapshot of the graph.
- We may observe the size of largest connected component, in generally, grow up over the time.

Evolution of the Scientific Communities

Three years window
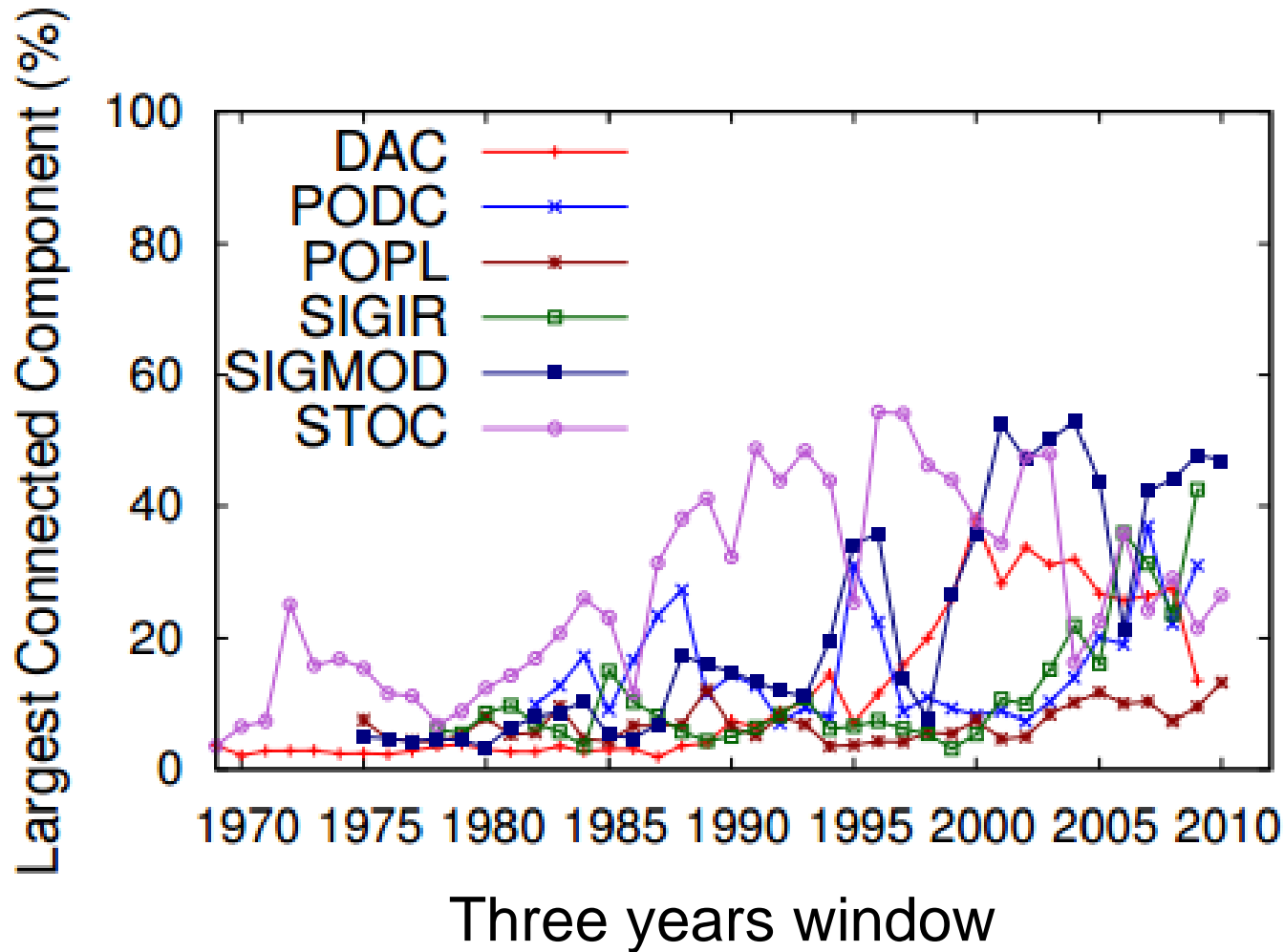
Snapshots (3 years) constructed based on nodes and edges

- The second perspective consists of analyzing snapshots constructed based on nodes and edges created on a predefined time window.

- Here, using the sliding window, we can note the variation over years.

- This motivated us to investigate the role of core community in this case.

Core Members vs. Non-Members
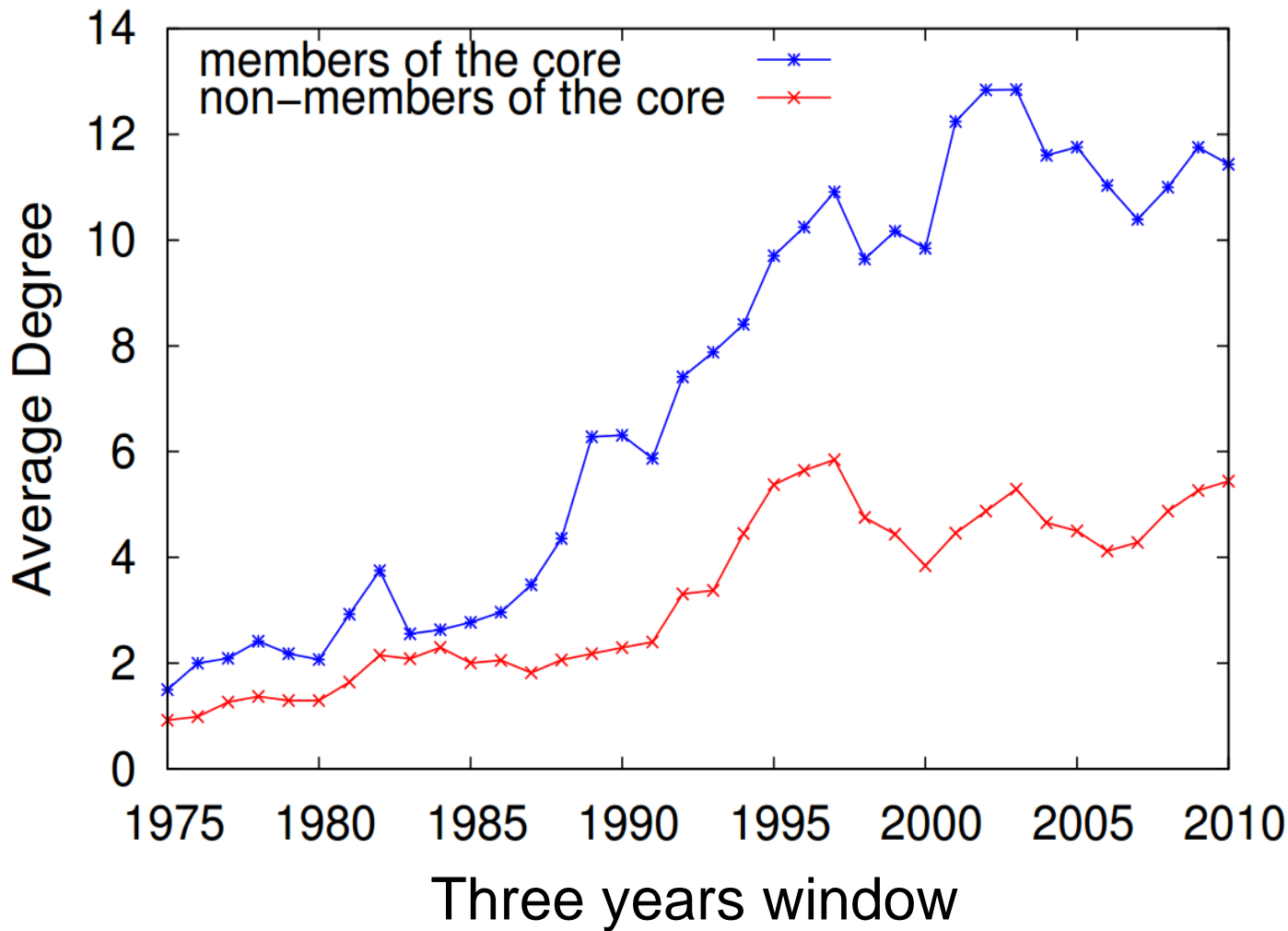
Average degree of core members is higher than those of non-members'

- The focus of this work is to understand the effects of the community core in the properties of the networks, so, using the sliding windows, we can perceive that average degree is higher for core members than non-members.
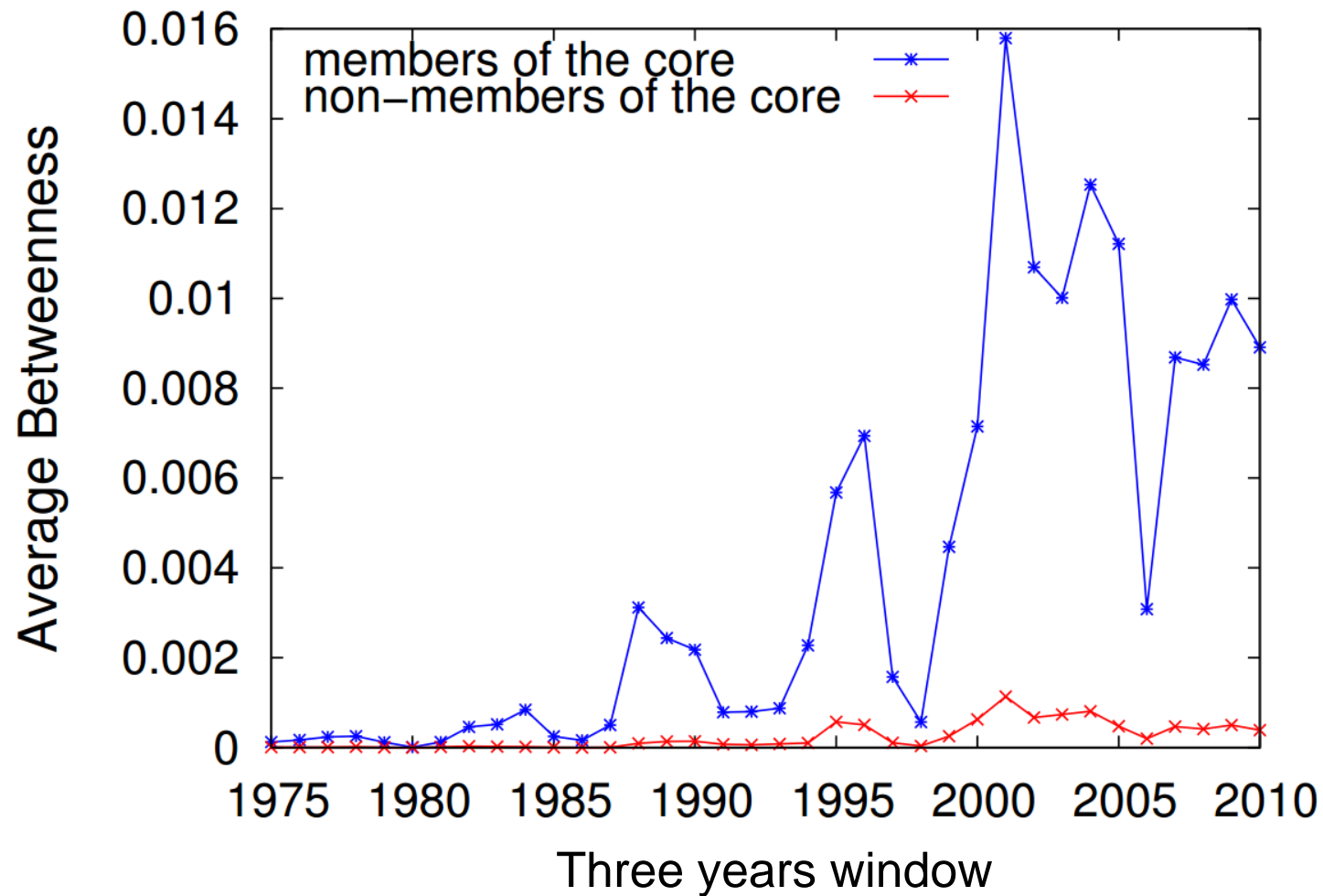
Core Members vs. Non-Members
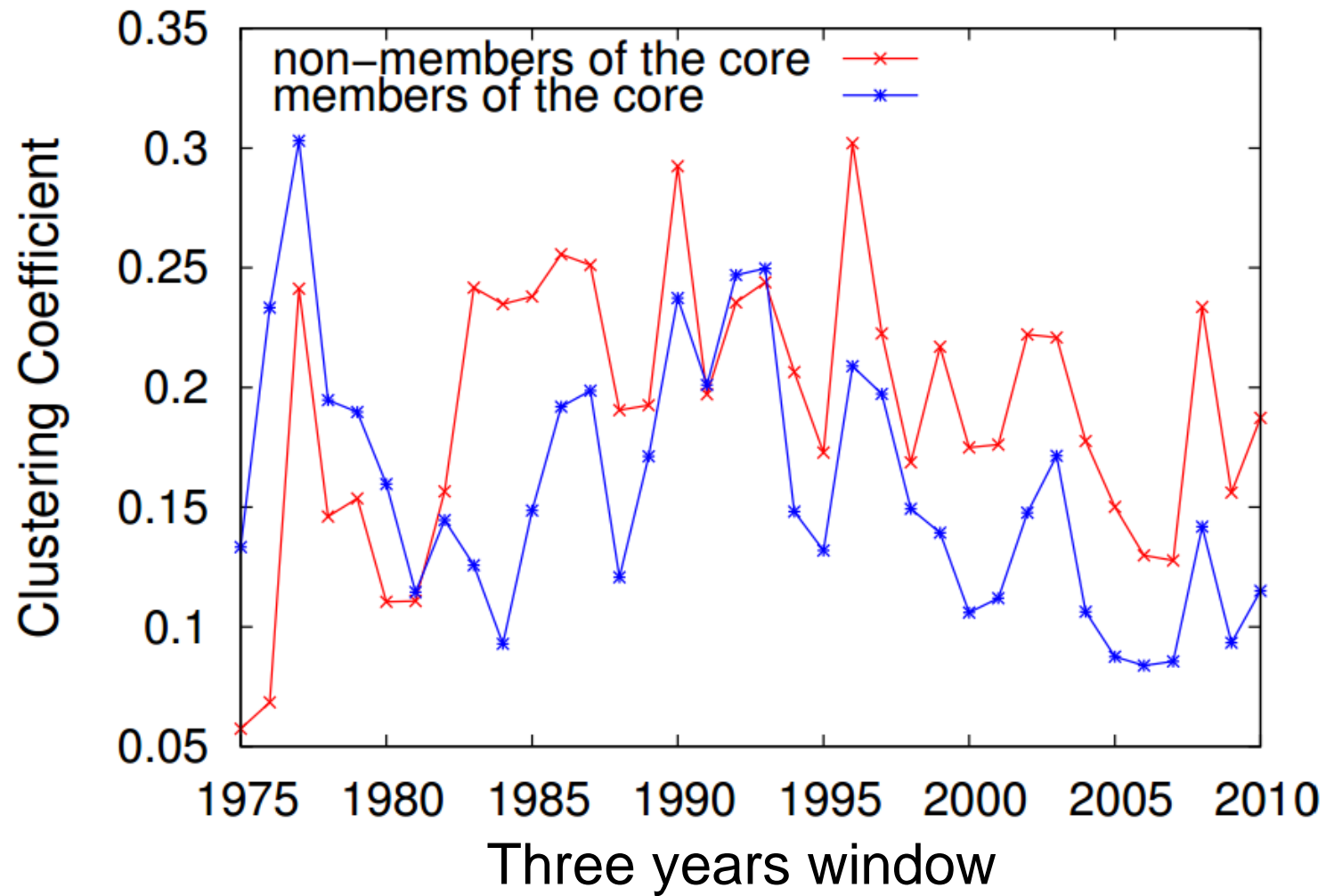
The core might act like hubs, by connecting different groups with small intersection

- Differently from the average degree, the clustering coefficient is smaller in community core, so the core might act like hubs in the network, by connecting different groups with small intersection.

Core Members vs. Non-Members

A higher number of shortest paths include the core

- So, to identify if the members of community core are hubs, we calculated the betweenness of the community core.

- The betweenness show which a higher number of shortest paths include the core. It confirms that members of community core are hubs.
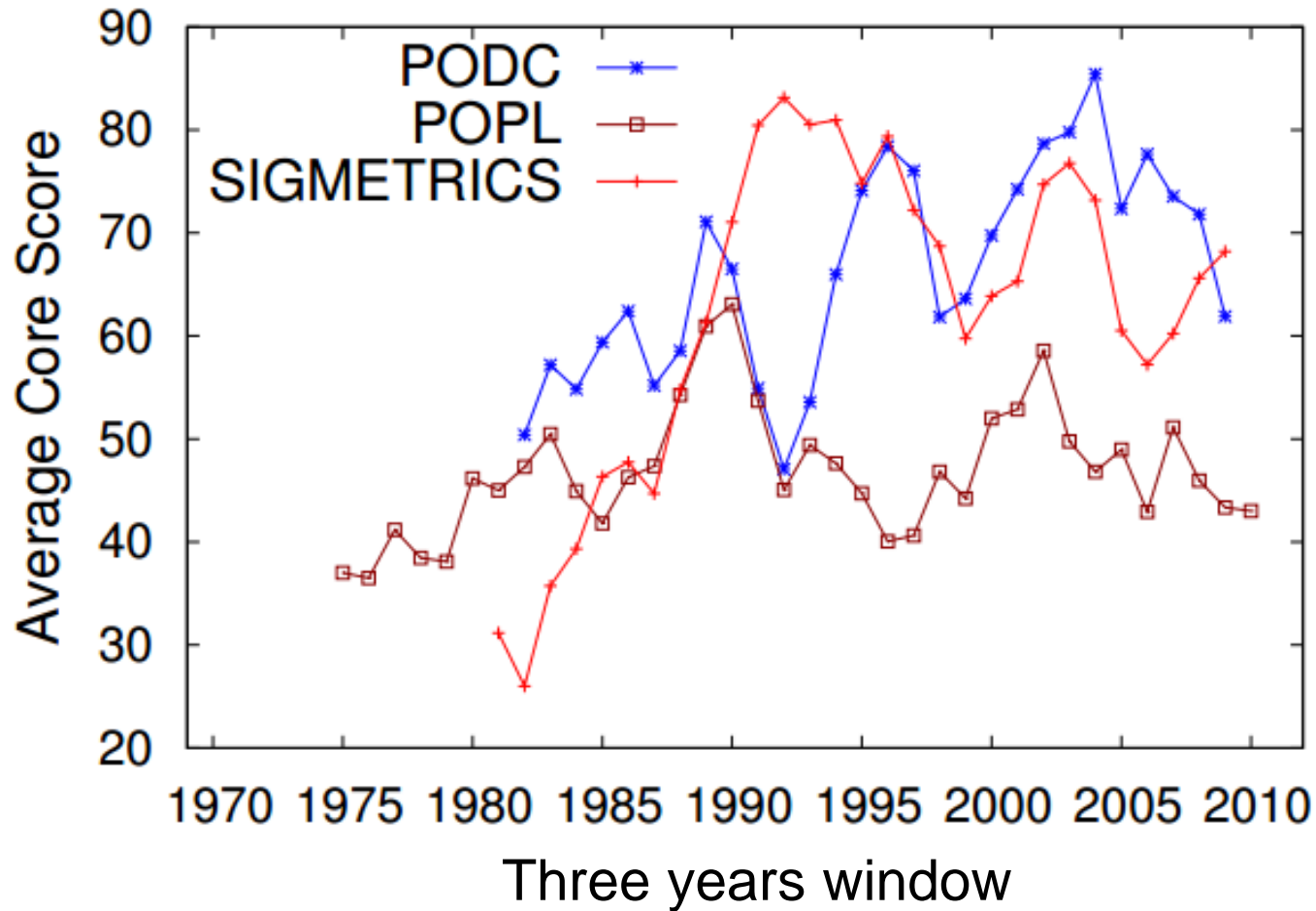
Influence of Core Members

Three years window
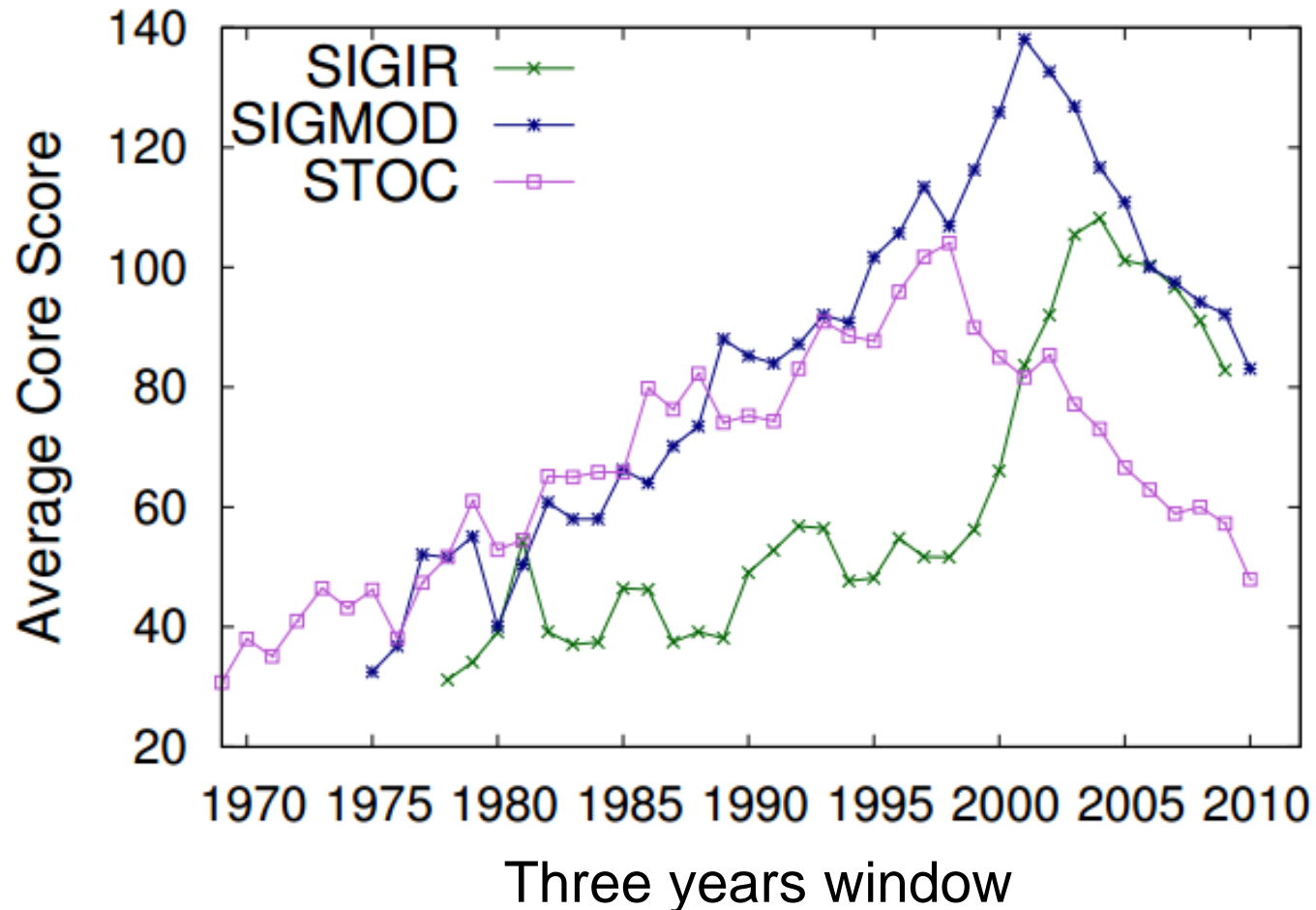
The average core score of a community, in general, rises along its life time

- To understand how the core score evolves over the time, we propose this metric, the average core score. The average core score of a community, in general, rises along its life time.
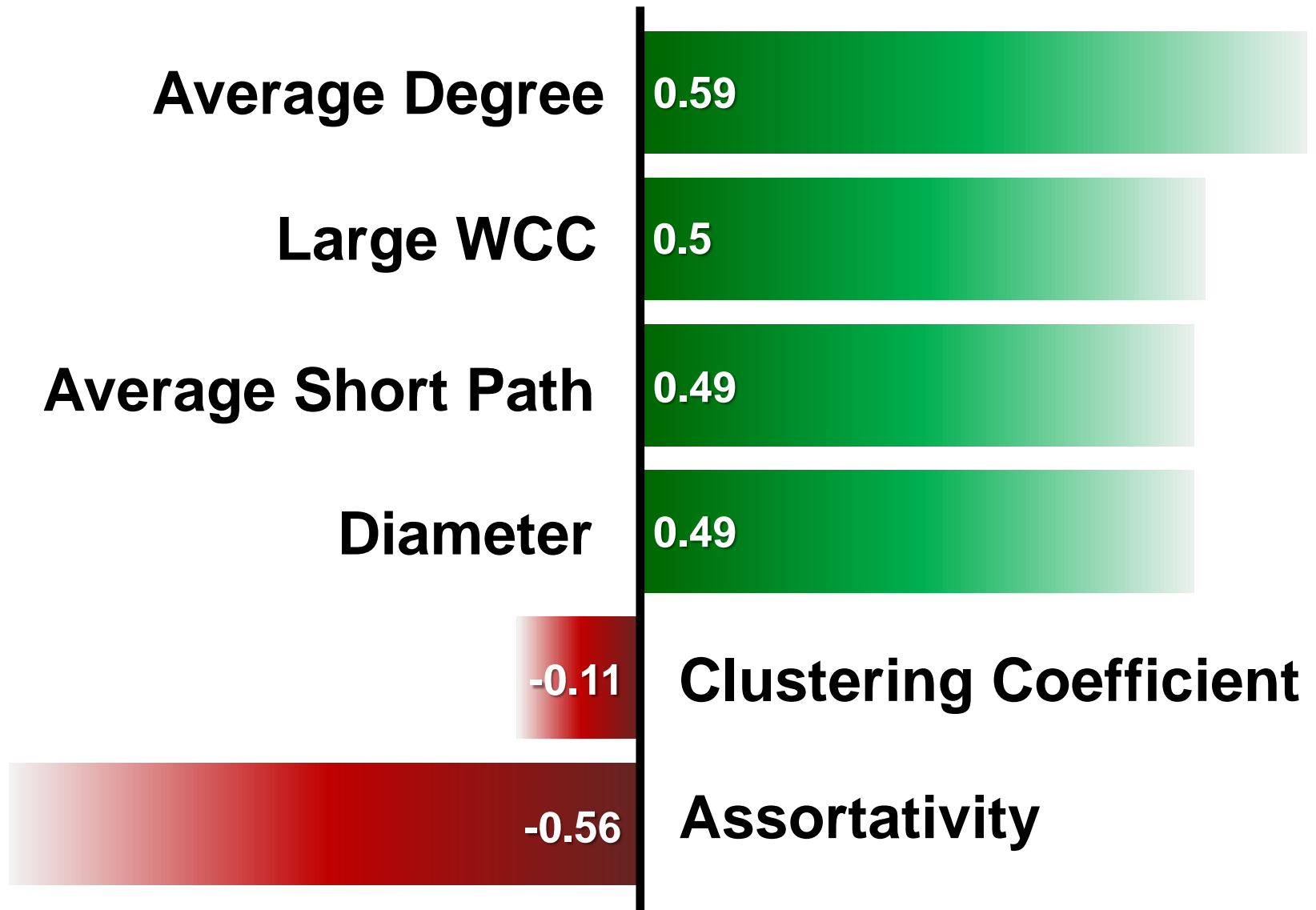
Influence of Core Members

Three years window

How does it affect the network?

- Here is possible observe a similar behavior, but How does it affect the network?

Strong Correlation with Average Core Score

Average Degree	0.59
Large WCC	0.5
Average Short Path	0.49
Diameter	0.49
-0.11	Clustering Coefficient
-0.56	Assortativity

- In this way, we calculated the correlation between the average core score and other complex networks metrics.

- Except for clustering coefficient, every metric has a strong correlation with average core score, in other words, if we increase or decrease the average core score, these metrics also change.

Conclusions

The core community:

- is strongly correlated with variations on network properties

- works as bridges that connect groups

- increases the average degree

- decreases the assortativeness

- Well, with this work, we conclude that:
- The core community work as bridges that connect smaller clustered research groups, and finally, the core community tends to increase the average degree of the network and decrease the assortativeness.

Conclusions

Future Work:

- Analysis of other kinds of network such as massive multiplayer games and on-line social networks

- Estimation of the h-index as a function of time

- As future work we propose to make analysis of other kinds of network such as massive multiplayer games and on-line social networks, and to estimate the h-index as a function of time.

Thank you!

Bruno Leite Alves

Fabrício Benevenuto   Alberto H. F. Laender

Computer Science Department

Federal University of Minas Gerais

{bruno.leite,fabricio,laender@dcc.ufmg.br}

- Thank you very much!