



# Aspectos Temporais para Medir a Força da Colaboração no GitHub

Natércia A. Batista, Michele A. Brandão,  
Ana Paula C. da Silva, Mirella M. Moro



**SBBD2017**

KDMILE

**BRACIS2017**

ENIAC

STIL

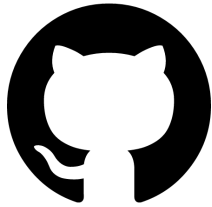
“

***Tarefa primordial da área de BD***

*Enriquecer dados existentes em  
busca de informações relevantes*

# Contexto

- ❑ **Social Coding** (desenvolv. colaborativo software)



- ❑ Novo tipo de **rede de colaboração**:  
uma colaboração ocorre quando dois desenvolvedores contribuem em um mesmo repositório

Desenvolv.  
Colaborativo  
de Software

Força de  
Colaboração

## Enriquecimento

### Caracterização

Rocha et al@SBBD16:  
Análise de contribuição  
de código entre  
repositórios do GitHub

Brandão & Moro @  
ComCom 2017: Social  
professional networks-  
a survey and taxonomy

### Métricas

Alves et al@SBBD16:  
The strength of social  
coding collaboration on  
GitHub [semântica](#)

Batista et al @ WI17:  
Collaboration strength  
metrics and analyses  
on GitHub [tempo](#)

## Informação Relevante

### NOW

>> Tutorial 3

>> [Batista et al @SBBD17](#)

[Janelas Temporais](#)

### NEXT

Recomendação de times

Predição de links

(planejamento, custos, etc)

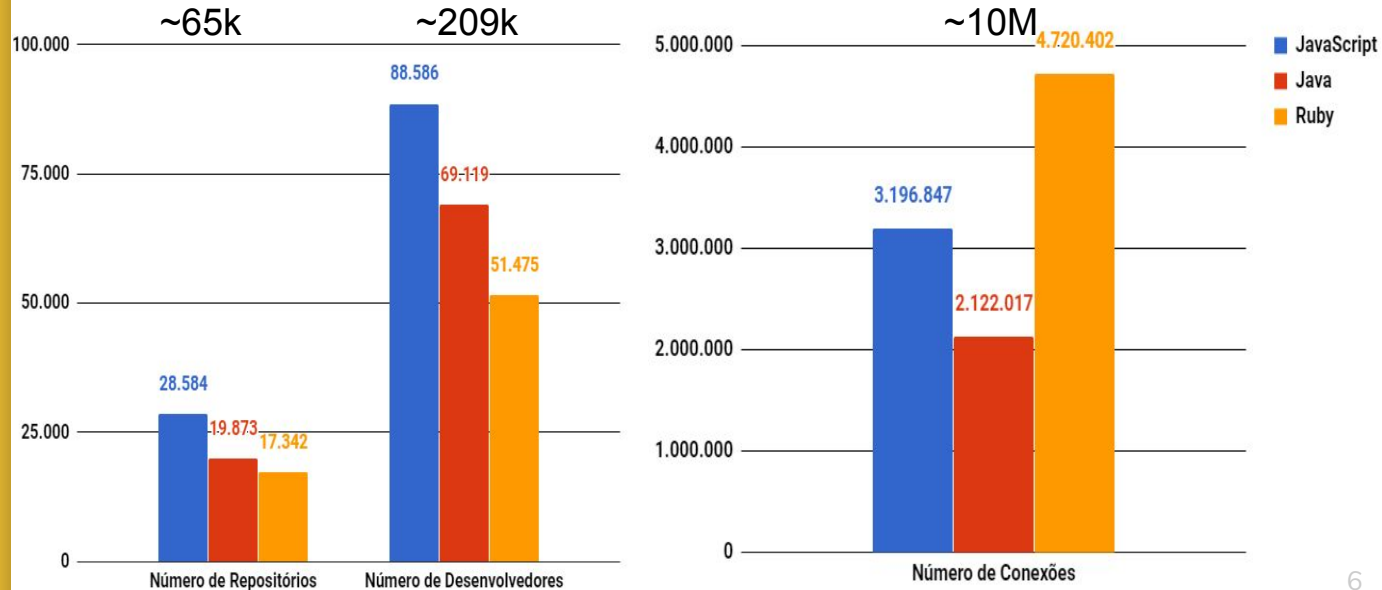
# Objetivos



1. Identificar **propriedades semânticas** provenientes do relacionamento entre desenvolvedores
2. Analisar a utilização das propriedades para definir a **força do relacionamento**
3. Melhorar a **modelagem da rede** com a utilização de dados temporais

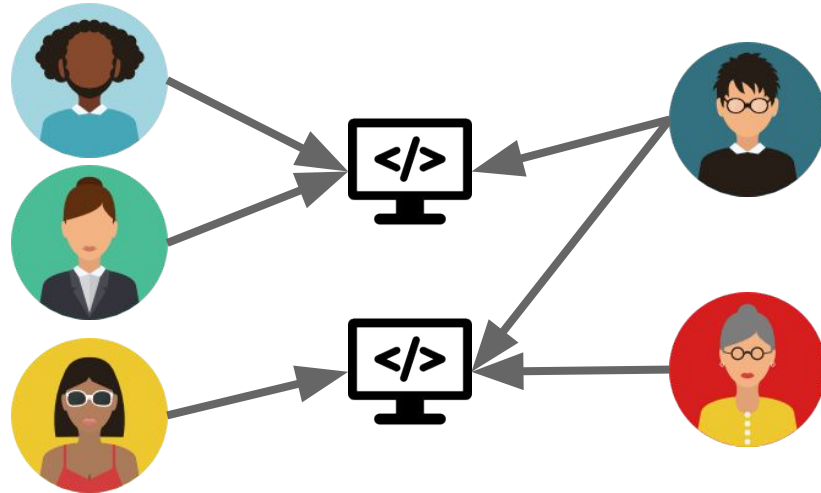
## GitSED (SBBD DSW 2017)

- 3 linguagens com o maior número de repositórios: **JavaScript**, **Java** e **Ruby**
- Repositórios base com 2+ desenvolvedores
- Excluídos repositórios forked e deleted

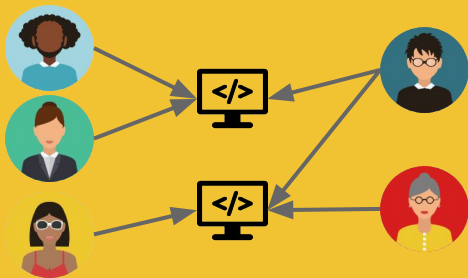
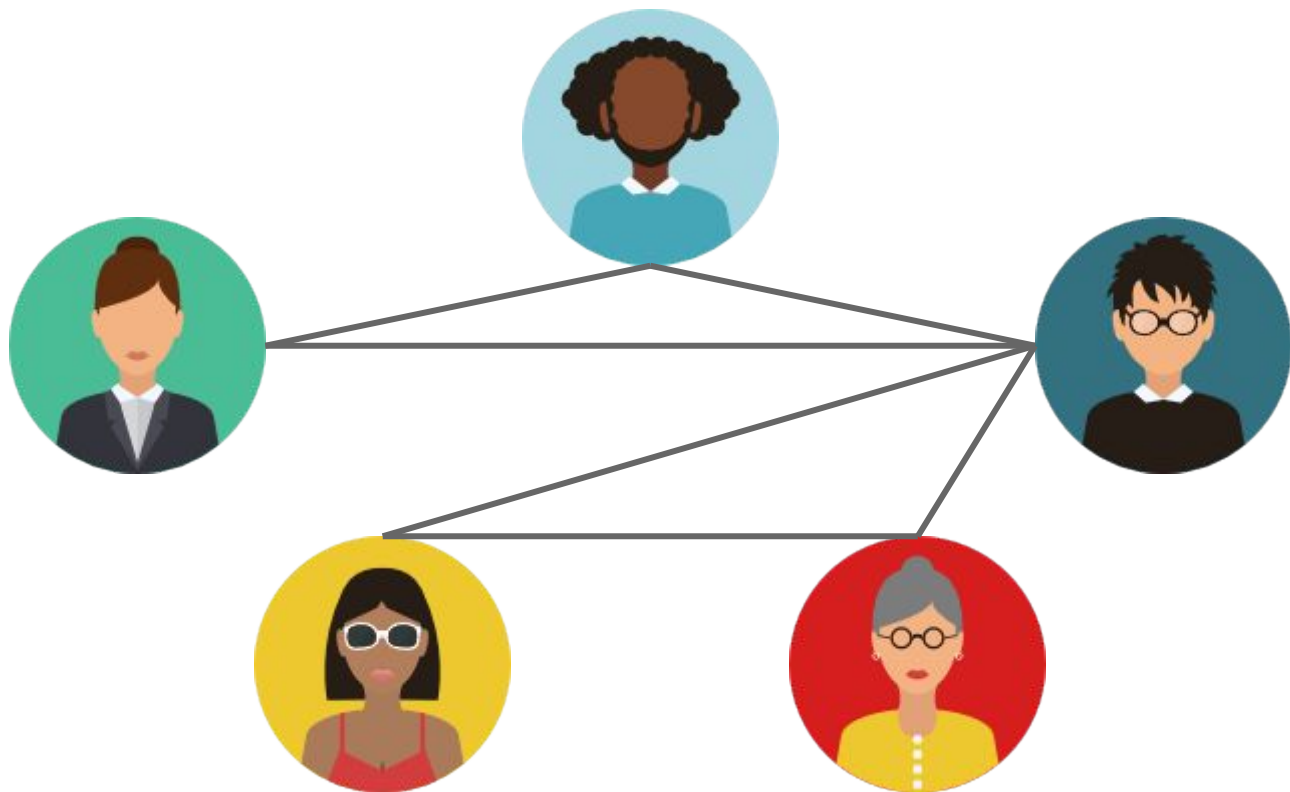


# Modelagem da Rede

- ❑ **Nós:** desenvolvedores
- ❑ **Arestas:** contribuição em repositório compartilhado
- ❑ **Peso:** métricas força de contribuição

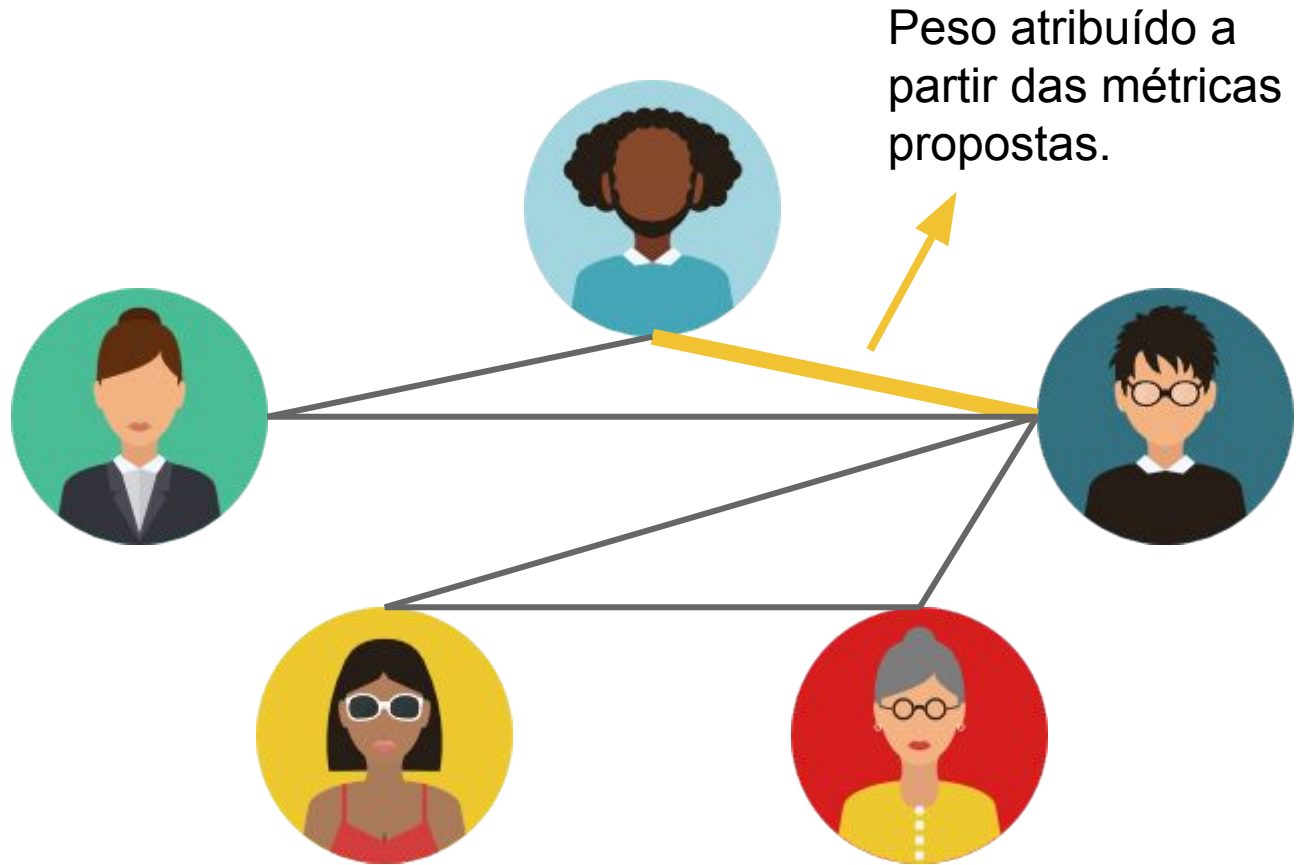
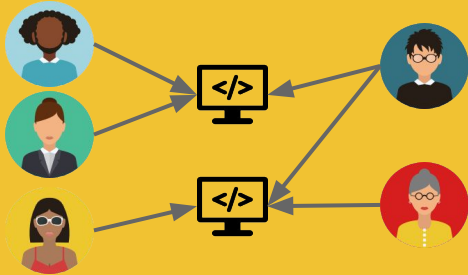


# Modelagem da Rede





# Modelagem da Rede



# Força do Relacionamento



## 1. Propriedades Topológicas



### Propriedades existentes

#### ❑ Neighborhood Overlap (NO)

Mede a força dos links (Easley and Kleinberg 2010)

#### ❑ Adamic-Adar Coefficient (AA)

Maior peso para vizinhos comuns com baixo grau

#### ❑ Preferential Attachment (PA)

Probabilidade de conexão maior de acordo com maior grau (Barabási and Albert 1999)

## 2. Propriedades Semânticas



### Propostas no SBBD 2016

- ❑ **Number of Shared Repositories (SR)**

Quantidade de repositórios compartilhados

- ❑ **Jointly developers contribution to shared repositories (JCSR)**

Contribuição de um par em relação ao total de desenvolvedores no repositório

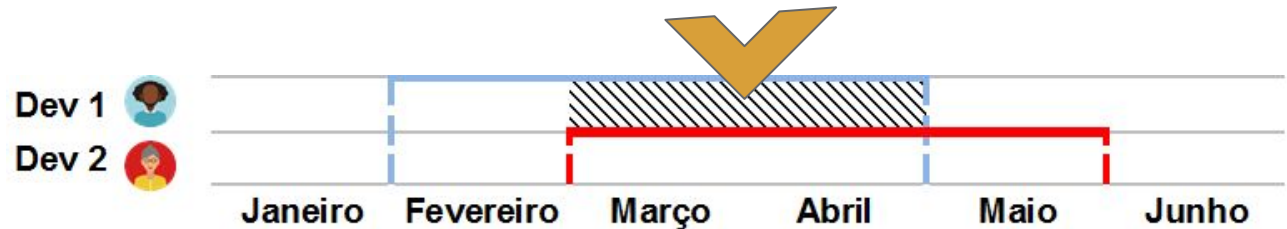
- ❑ **Jointly developers commits to shared repositories (JCOSR)**

Quantidade de commits de um par de desenvolvedores

### 3. Propriedades Semânticas com Aspecto Temporal

#### Local Potential Contribution (LPC)

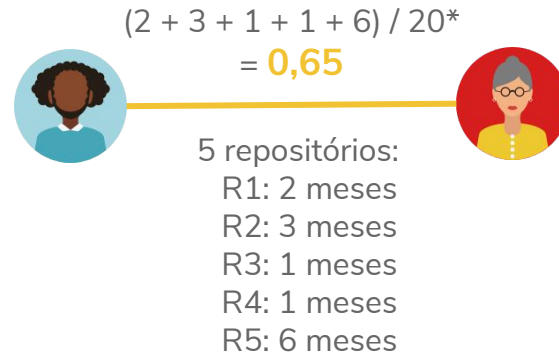
Tempo de contribuição do par de desenvolvedores  
em repositórios compartilhados



### 3. Propriedades Semânticas com Aspecto Temporal

#### Global Potential Contribution (GPC)

Tempo de contribuição do par de desenvolvedores em relação ao maior tempo de contribuição da rede

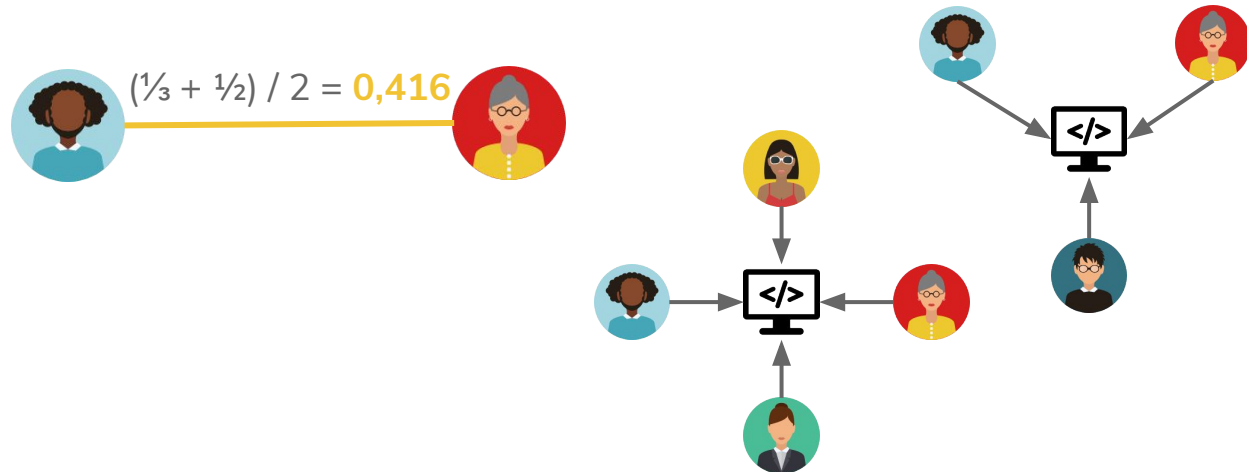


\* maior tempo de contribuição na rede

### 3. Propriedades Semânticas com Aspecto Temporal

#### Previous Collaboration (PC)

Contribuição em repositórios passados em relação ao número de desenvolvedores nos mesmos



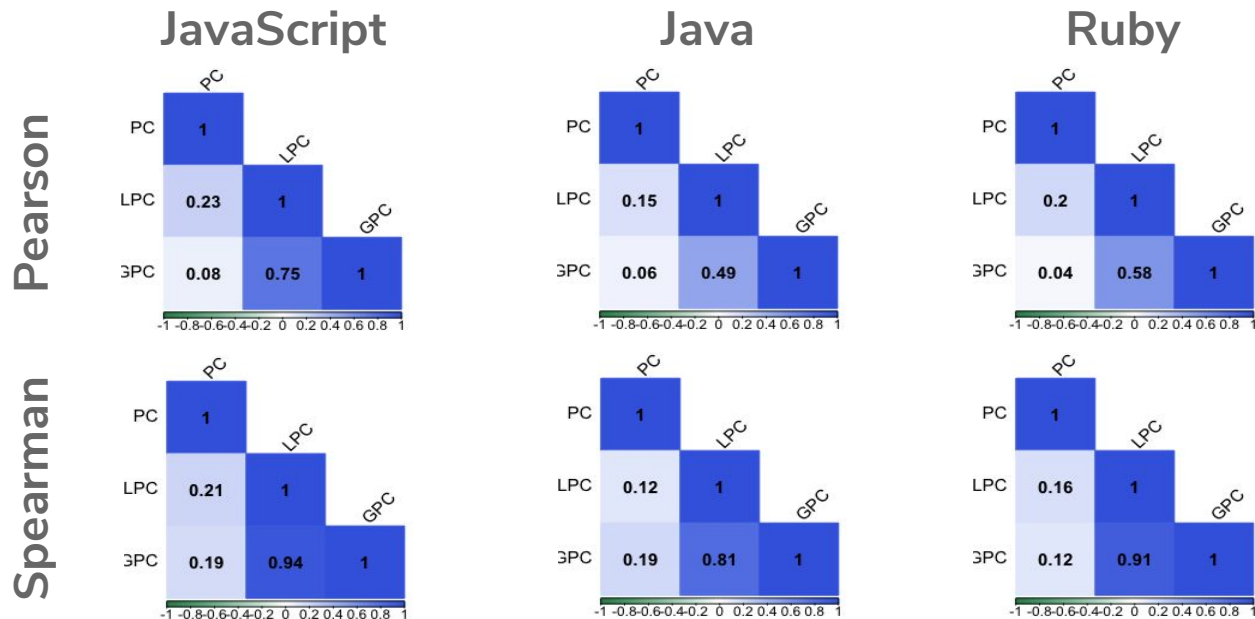
# Força do Relacionamento

## 4. Correlação entre Propriedades

Correlação entre propriedades semânticas com aspecto temporal

(1) LPC e GPC: correlacionadas para todas as linguagens

(2) PC não correlacionada devido à interseção do tempo de colaboração (mesma janela de tempo)





## N imagens no artigo

### Todas as propriedades

- ❑ O tempo de contribuição (LPC e GPC) bem relacionado às métricas JCSR e JCOSR
- ❑ Entre as métricas topológicas, devem ser consideradas NO e apenas uma entre PA e AA (altamente correlacionadas)

JCSR - Jointly developers contribution to shared repositories

JCOSR - Jointly developers commits to shared repositories

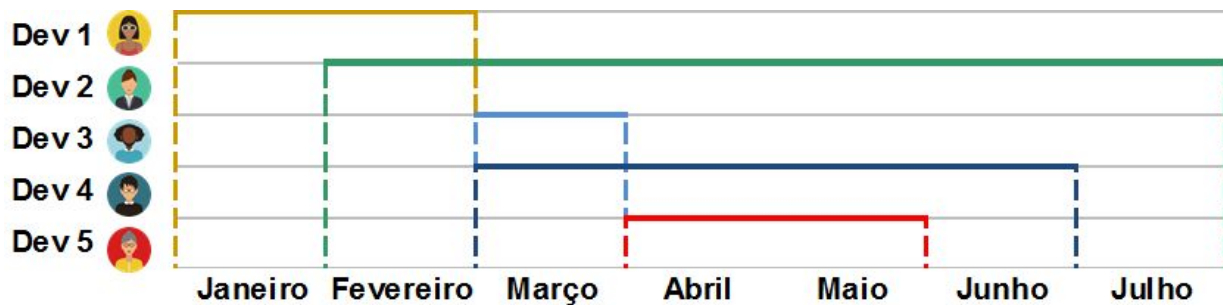
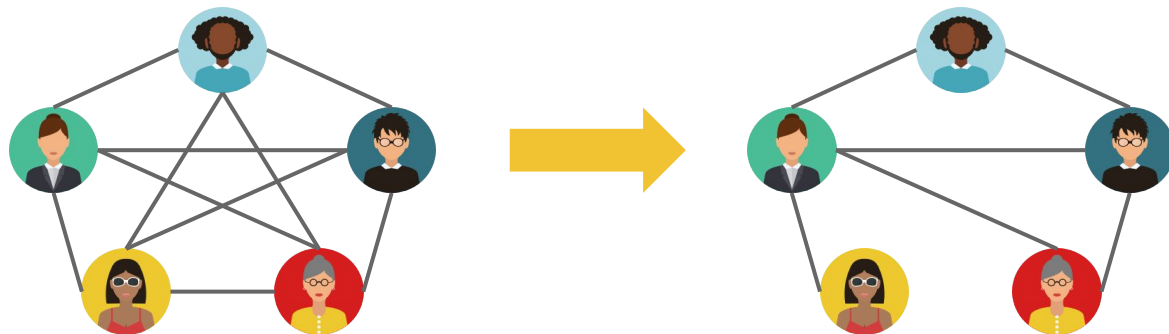
NO - Neighborhood Overlap

AA - Adamic-Adar Coefficient

PA - Preferential Attachment

# \*\*\*Nova\*\*\* Modelagem Proposta

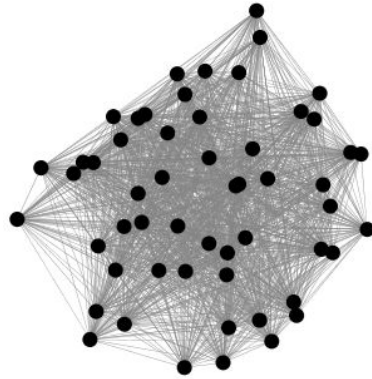
A maior parte dos desenvolvedores atua num repositório em **janelas de tempo** diferentes



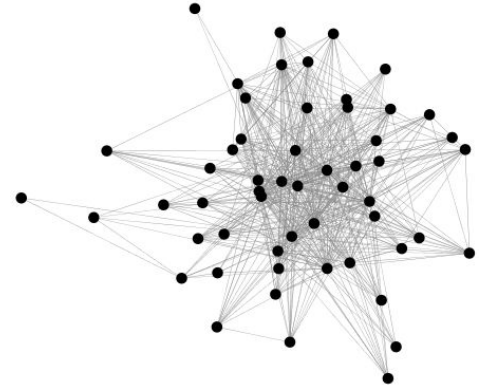
# Nova Modelagem Proposta

Exemplo real da rede:

- ❑ Repositório JavaScript com 55 nós
- ❑ Duração: março/2012 a setembro/2015



- ❑ Arestas: 1.485
- ❑ Densidade: 1,000
- ❑ Grau Médio: 54,0



- ❑ Arestas: 571
- ❑ Densidade: 0,385
- ❑ Grau Médio: 20,8

# Trabalhos Futuros



1. Incrementar as análises com outras linguagens de programação  
**GitSED @ SBBD DSW 2017**
2. Selecionar linguagens a partir do seu grau de relacionamento (alto e baixo)
3. Aplicar a nova modelagem atribuindo arestas a partir da contribuição na mesma janela de tempo



**Obrigada!**

**Dúvidas?**

- [natercia@dcc.ufmg.br](mailto:natercia@dcc.ufmg.br)
- [micheleabrandao@dcc.ufmg.br](mailto:micheleabrandao@dcc.ufmg.br)
- [ana.coutosilva@dcc.ufmg.br](mailto:ana.coutosilva@dcc.ufmg.br)
- [mirella@dcc.ufmg.br](mailto:mirella@dcc.ufmg.br)

# Aspectos Temporais para Medir a Força da Colaboração no GitHub

Natércia A. Batista, Michele A. Brandão,  
Ana Paula C. da Silva, Mirella M. Moro



## Trabalhos Relacionados

- ❑ WI 2017: resultados iniciais \*sem\* a noção de janelas de tempo
- ❑ Métricas para colaboração entre desenvolvedores a partir de dados relacionados ao **repositório**: repositórios compartilhados ou quantidade de commits dos desenvolvedores (Alves, et al. 2016)
- ❑ Novas métricas: **relacionamentos anteriores** (Casalnuovo et al. 2015) e **dados temporais** das interações

# LPC x GPC

Local Potential Contribution vs. Global Potential Contribution

Par Desenv.	Repositório	Tempo Contrib.	Duração	LPC
A, B	R1	3 meses	3 meses	$3/3 = 1,00$
C, D	R2	6 meses	12 meses	$6/12 = 0,50$

Par Desenv.	Tempo Contrib.	Maior Tempo Contrib. (rede inteira)	GPC
A, B	3 meses	12 meses	$3/12 = 0,25$
C, D	6 meses		$6/12 = 0,50$