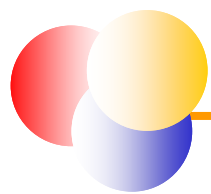


计算机数学建模

第二讲 简单优化模型(1)

周毓明

南京大学计算机科学与技术系



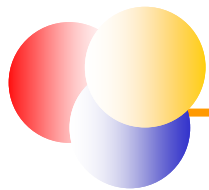
课程内容

1. 量纲分析
2. Google的“秘诀”



1. 量纲分析





量纲分析

问题：单摆运动

求摆动周期 t 的表达式

摆的角速度为 θ' ，角加速度为 θ''

由牛顿第二运动定律，有

$$(m) * (l * \theta'') = -mg * \sin\theta$$

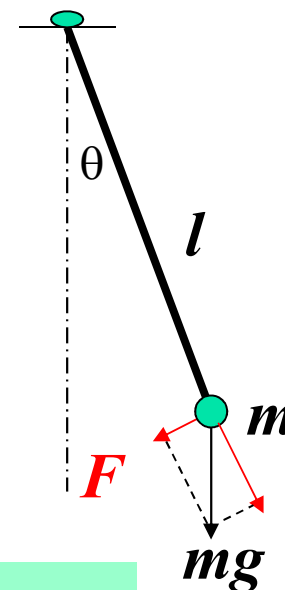
$$\text{即 } \theta'' + (g/l) * \sin\theta = 0$$

令 $\omega = (g/l)^{1/2}$ ，有

$$\theta'' + (\omega^2) * \sin\theta = 0$$

当 θ 很小时， $\sin\theta \approx \theta$

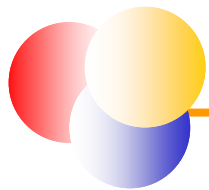
$$F = mg \sin\theta$$



$$\theta'' + (\omega^2) * \theta \approx 0$$

该方程的解为 $\theta = A * \sin(\omega t + \phi)$

$$t = 2\pi / \omega = 2\pi \sqrt{\frac{l}{g}}$$



量纲分析

量纲齐次原则

物理量的量纲

长度 l 的量纲记 $L=[l]$

质量 m 的量纲记 $M=[m]$

时间 t 的量纲记 $T=[t]$

速度 v 的量纲 $[v]=LT^{-1}$

加速度 a 的量纲 $[a]=LT^{-2}$

力 f 的量纲 $[f]=LMT^{-2}$

引力常数 k 的量纲 $[k] = [f][l]^2[m]^{-2} = L^3M^{-1}T^{-2}$

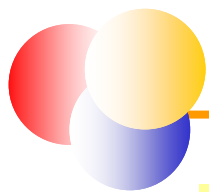
对无量纲量 α , $[\alpha]=1(=L^0M^0T^0)$ $f = k \frac{m_1 m_2}{r^2}$



动力学中
基本量纲
 L, M, T



导出量纲



量纲分析

量纲齐次原则

等式两端的量纲一致

量纲分析~利用量纲齐次原则寻求物理量之间关系

例：单摆运动

求摆动周期 t 的表达式

设物理量 t, m, l, g
之间有关系式

$$t = \lambda m^{\alpha_1} l^{\alpha_2} g^{\alpha_3} \quad (1)$$

$\alpha_1, \alpha_2, \alpha_3$ 为待定系数, λ 为无量纲量

(1) 的量纲表达式

$$[t] = [m]^{\alpha_1} [l]^{\alpha_2} [g]^{\alpha_3}$$

$$\Rightarrow T = M^{\alpha_1} L^{\alpha_2 + \alpha_3} T^{-2\alpha_3}$$

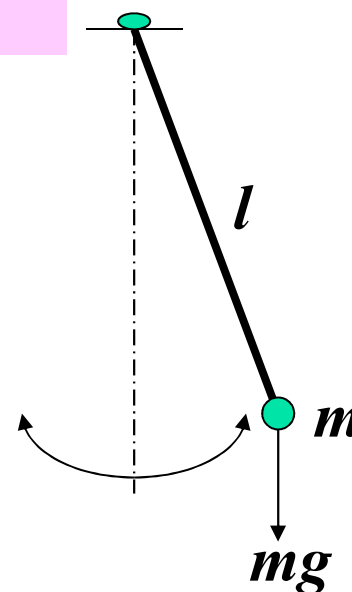
$$\Rightarrow \begin{cases} \alpha_1 = 0 \\ \alpha_2 + \alpha_3 = 0 \\ -2\alpha_3 = 1 \end{cases}$$

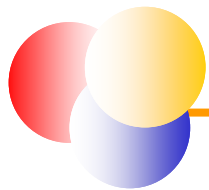
$$\Rightarrow \begin{cases} \alpha_1 = 0 \\ \alpha_2 = 1/2 \\ \alpha_3 = -1/2 \end{cases}$$

$$\Rightarrow t = \lambda \sqrt{\frac{l}{g}}$$

对比

$$t = 2\pi \sqrt{\frac{l}{g}}$$





量纲分析

$$t = \lambda m^{\alpha_1} l^{\alpha_2} g^{\alpha_3}$$

为什么假设这种形式

设 $p = f(x, y, z)$

对 x, y, z 的两组测量值 x_1, y_1, z_1 和 x_2, y_2, z_2 ,
 $p_1 = f(x_1, y_1, z_1)$, $p_2 = f(x_2, y_2, z_2)$

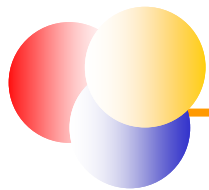
x, y, z 的量纲单位
缩小 a, b, c 倍

$$p'_1 = f(ax_1, by_1, cz_1), p'_2 = f(ax_2, by_2, cz_2)$$

$$\frac{p_1}{p_2} = \frac{p'_1}{p'_2} \quad \Rightarrow \quad \frac{f(x_1, y_1, z_1)}{f(x_2, y_2, z_2)} = \frac{f(ax_1, by_1, cz_1)}{f(ax_2, by_2, cz_2)}$$

\Rightarrow $p = f(x, y, z)$ 的形式为 $f(x, y, z) = \lambda x^\alpha y^\beta z^\gamma$





量纲分析

单摆运动中 t, m, l, g 的一般表达式

$$f(t, m, l, g) = 0$$

$$\Rightarrow t^{y_1} m^{y_2} l^{y_3} g^{y_4} = \pi \quad y_1 \sim y_4 \text{ 为待定常数, } \pi \text{ 为无量纲量}$$

$$\begin{cases} [t] = L^0 M^0 T^1 \\ [m] = L^0 M^1 T^0 \\ [l] = L^1 M^0 T^0 \\ [g] = L^1 M^0 T^{-2} \end{cases}$$

$$(L^0 M^0 T^1)^{y_1} (L^0 M^1 T^0)^{y_2} (L^1 M^0 T^0)^{y_3}$$

$$(L^1 M^0 T^{-2})^{y_4} = L^0 M^0 T^0$$

$$L^{y_3+y_4} M^{y_2} T^{y_1-2y_4} = L^0 M^0 T^0$$

$$\begin{cases} y_3 + y_4 = 0 \\ y_2 = 0 \\ y_1 - 2y_4 = 0 \end{cases}$$

基本解 y

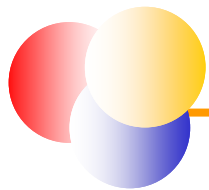
$$= (y_1, y_2, y_3, y_4)^T$$

$$= (2, 0, -1, 1)^T$$

$$t^2 l^{-1} g = \pi$$

$$F(\pi) = 0$$

$$(t = \lambda \sqrt{l/g})$$



量纲分析

Pi定理 (Buckingham)

设 $f(q_1, q_2, \dots, q_m) = 0$

是与量纲单位无关的物理定律, X_1, X_2, \dots, X_n 是基本量纲, $n \leq m$, q_1, q_2, \dots, q_m 的量纲可表为

$$[q_j] = \prod_{i=1}^n X_i^{a_{ij}}, \quad j = 1, 2, \dots, m$$

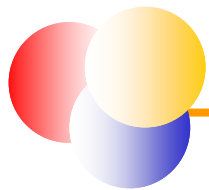
量纲矩阵记作 $A = \{a_{ij}\}_{n \times m}$, 若 $\text{rank } A = r$

线性齐次方程组 $Ay = 0$ 有 $m-r$ 个基本解, 记作

$$y_s = (y_{s1}, y_{s2}, \dots, y_{sm})^T, \quad s = 1, 2, \dots, m-r$$

则 $\pi_s = \prod_{j=1}^m q_j^{y_{sj}}$ 为 $m-r$ 个相互独立的无量纲量, 且

$F(\pi_1, \pi_2, \dots, \pi_{m-r}) = 0$ 与 $f(q_1, q_2, \dots, q_m) = 0$ 等价, F 未定



量纲分析

量纲分析示例：波浪对航船的阻力

航船阻力 f

航船速度 v , 船长 l , 吃水深度 h , 海水密度 ρ , 粘性系数 μ , 重力加速度 g 。

$$f(q_1, q_2, \dots, q_m) = 0$$

$$\phi(f, l, h, v, \rho, \mu, g) = 0$$

$$[q_j] = \prod_{i=1}^n X_i^{a_{ij}}, \\ j = 1, 2, \dots, m$$

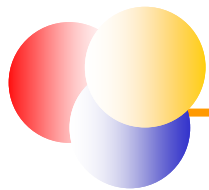
$$A = \{ a_{ij} \}_{n \times m}$$

$$m=7, n=3$$

$$[f] = LMT^{-2}, [l] = L, [h] = L, [v] = LT^{-1}, \\ [\rho] = L^{-3}M, [\mu] = L^{-1}MT^{-1}, [g] = LT^{-2},$$

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & -3 & -1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ -2 & 0 & 0 & -1 & 0 & -1 & -2 \end{bmatrix} \begin{matrix} (L) \\ (M) \\ (T) \end{matrix}$$

(f) (l) (h) (v) (ρ) (μ) (g)



量纲分析

$$f(q_1, q_2, \dots, q_m) = 0$$

$$\text{rank } A = r$$

$Ay=0$ 有 $m-r$ 个基本解

$$y_s = (y_{s1}, y_{s2}, \dots, y_{sm})^T$$

$$s = 1, 2, \dots, m-r$$

$m-r$ 个无量纲量

$$\pi_s = \prod_{j=1}^m q_j^{y_{sj}}$$

$$\phi(f, l, h, v, \rho, \mu, g) = 0$$

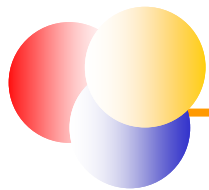
$$\text{rank } A = 3$$

$Ay=0$ 有 $m-r=4$ 个基本解

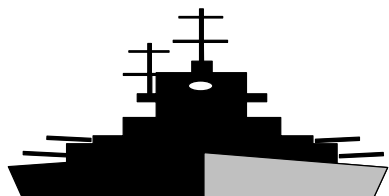
$$\begin{cases} y_1 = (0 & 1 & -1 & 0 & 0 & 0 & 0)^T \\ y_2 = (0 & 1 & 0 & -2 & 0 & 0 & 1)^T \\ y_3 = (0 & 1 & 0 & 1 & 1 & -1 & 0)^T \\ y_4 = (1 & -2 & 0 & -2 & -1 & 0 & 0)^T \end{cases}$$

$$\begin{cases} \pi_1 = lh^{-1} \\ \pi_2 = lv^{-2} \\ \pi_3 = lv\rho\mu^{-1} \\ \pi_4 = fl^{-2}v^{-2}\rho^{-1} \end{cases}$$





量纲分析



$F(\pi_1, \pi_2, \dots, \pi_{m-r}) = 0$ 与
 $f(q_1, q_2, \dots, q_m) = 0$ 等价

$$\pi_s = \prod_{j=1}^m q_j^{y_{sj}}$$

为得到阻力 f 的显式表达式

$F(\pi_1, \pi_2, \pi_3, \pi_4) = 0$ 与
 $\varphi(f, l, h, v, \rho, \mu, g) = 0$ 等价

$$\begin{cases} \pi_1 = lh^{-1} \\ \pi_2 = lv^{-2}g \\ \pi_3 = lv\rho\mu^{-1} \\ \pi_4 = fl^{-2}v^{-2}\rho^{-1} \end{cases}$$

$$F=0 \quad \Rightarrow \pi_4 = \psi(\pi_1, \pi_2, \pi_3)$$

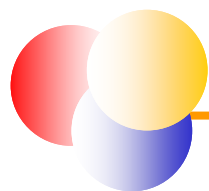


$$f = l^2 v^2 \rho \psi(\pi_1, \pi_2, \pi_3)$$



ψ 未定





量纲分析

量纲分析法的评注



- 物理量的选取

$\varphi(\dots) = 0$ 中包括哪些物理量是至关重要的

- 基本量纲的选取

基本量纲个数 n ; 选哪些基本量纲

- 基本解的构造

有目的地构造 $A\mathbf{y} = 0$ 的基本解

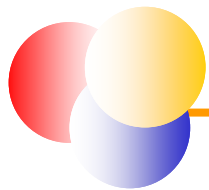
- 方法的普适性

不需要特定的专业知识

- 结果的局限性

函数 F 和无量纲量未定





量纲分析

量纲分析在物理模拟中的应用

例：航船阻力的物理模拟

通过航船模型确定原型船所受阻力

已知模
型船所
受阻力

$$f = l^2 v^2 \rho \psi(\pi_1, \pi_2)$$

$$\pi_1 = \frac{l}{h}, \pi_2 = \frac{lg}{v^2}$$

可得原
型船所
受阻力

$$f' = l'^2 v'^2 \rho' \psi(\pi'_1, \pi'_2)$$

$$\pi'_1 = \frac{l'}{h'}, \pi'_2 = \frac{l'g'}{v'^2}$$

$$f, l, h, v, \rho, g$$

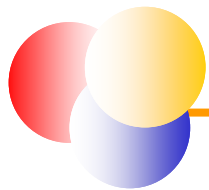
~模型船的参数(均已知)

$$f', l', h', v', \rho', g'$$

~原型船的参数

(f' 未知, 其他已知)

注意：二者的 ψ 相同

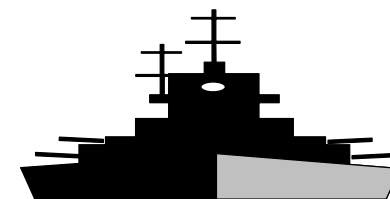


量纲分析

$$f = l^2 v^2 \rho \psi(\pi_1, \pi_2) \quad f' = l'^2 v'^2 \rho' \psi(\pi'_1, \pi'_2)$$

$$\pi_1 = \frac{l}{h}, \pi_2 = \frac{lg}{v^2}$$

$$\pi'_1 = \frac{l'}{h'}, \pi'_2 = \frac{l'g'}{v'^2}$$

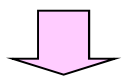


$$g = g'$$

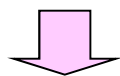
$$\pi_1 = \pi'_1,$$

$$\pi_2 = \pi'_2 \Rightarrow$$

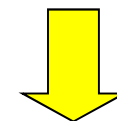
$$\frac{f'}{f} = \left(\frac{l'v'}{lv} \right)^2 \frac{\rho'}{\rho}$$



$$\frac{l}{h} = \frac{l'}{h'}$$



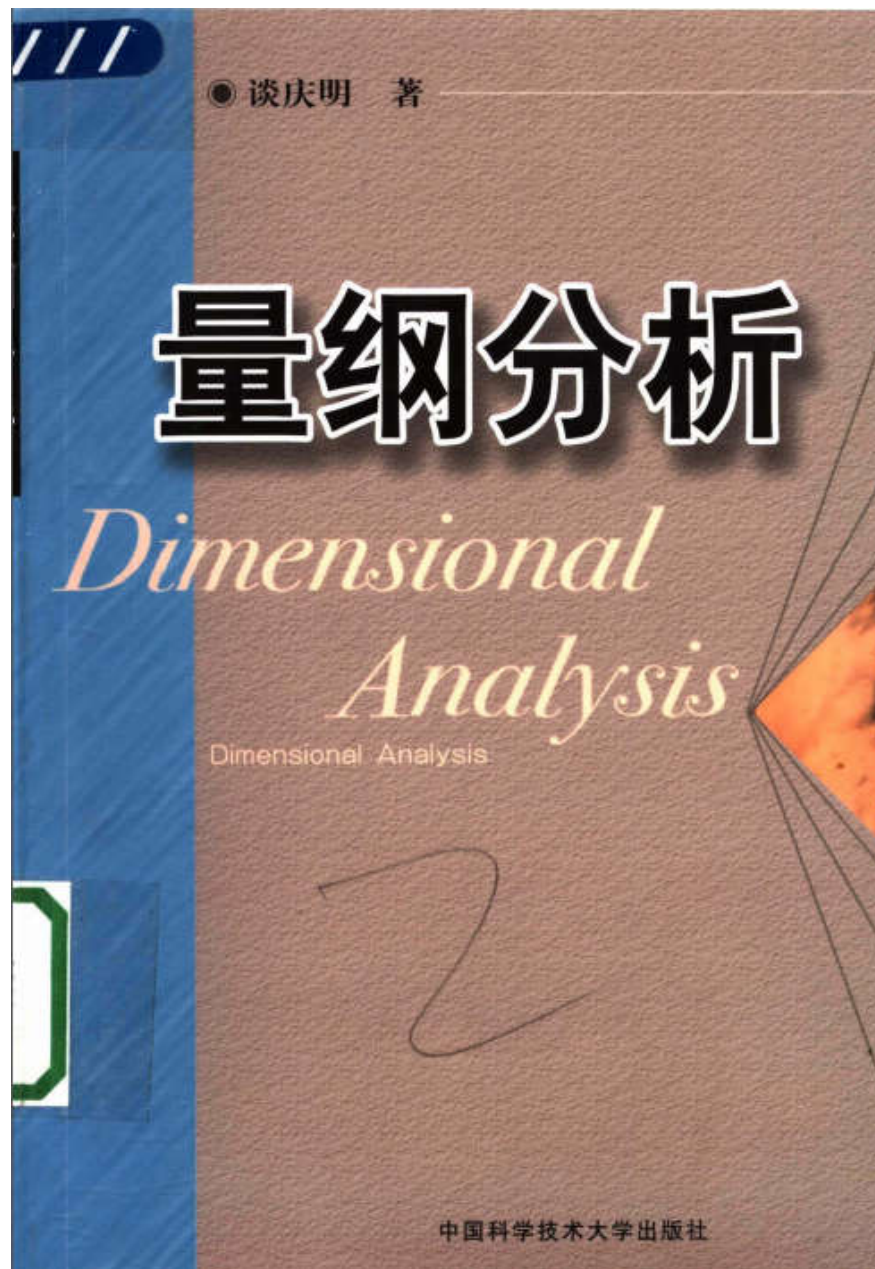
$$\frac{v}{v'} = \sqrt{\frac{lg}{l'g'}}$$



$$(\rho = \rho')$$

$$\frac{f'}{f} = \left(\frac{l'}{l} \right)^3$$

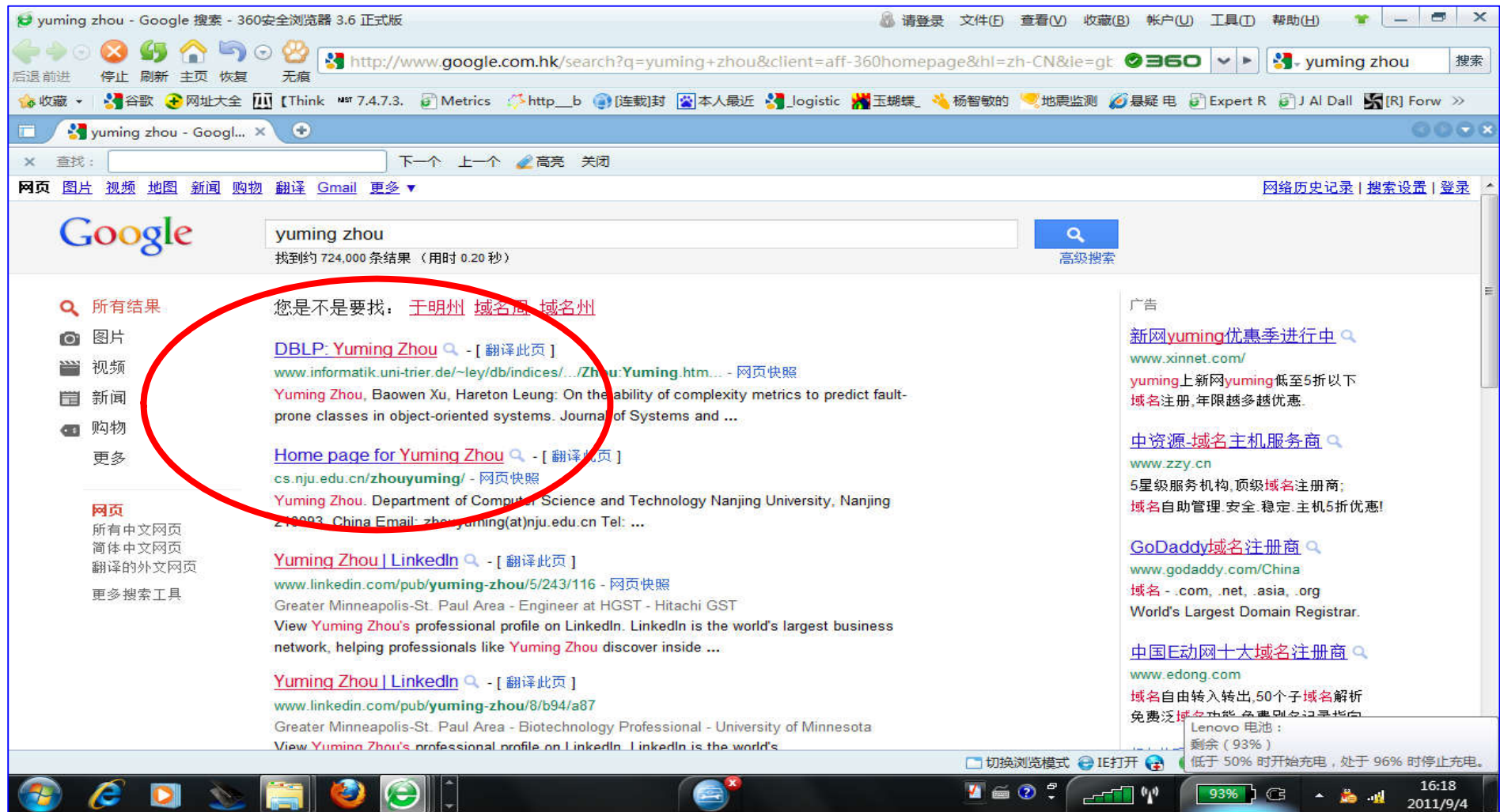
按一定尺寸比例造模型船，
量测 f ，可算出 原型 f'



2. Google的“秘诀”



2017年9月



2018年9月

文件(E) 编辑(E) 查看(V) 历史(S) 书签(B) 工具(T) 帮助(H)

Yuming Zhou - Google 搜索 X

https://www.google.com.hk/search?safe=strict&source=hp&ei=QReXW93WHaSAk-4PoOyEuAM&q=Yuming

火狐官方网站 最常访问 火狐官方网站 新手上路 常用网址 JD 京东商城 常用网址 JD 京东商城 移动版书签

Google Yuming Zhou

全部 图片 新闻 视频 地图 更多 设置 工具 已启用安全搜索

找到约 273,000 条结果 (用时 0.50 秒)

您是不是要找:

于明州 余明州 于明咒

dblp: Yuming Zhou
<https://dblp.uni-trier.de/Persons> 翻译此页
List of computer science publications by Yuming Zhou.

Home page for Yuming Zhou
<https://cs.nju.edu.cn/zhoudyuming/> 翻译此页
I am currently a professor in the Department of Computer Science and Technology at Nanjing University.
I received my Ph.D. degree in Computer Science from ...

Yuming Zhou - Director - WD, a Western Digital Company | LinkedIn
<https://www.linkedin.com/in/yuming-zhou-1162435> 翻译此页
View Yuming Zhou's profile on LinkedIn, the world's largest professional community. Yuming has 4 jobs listed on their profile. See the complete profile on ...

Top 25 Yuming Zhou profiles | LinkedIn
<https://www.linkedin.com/pub/dir/Yuming/Zhou>
View the profiles of professionals named Yuming Zhou on LinkedIn. There are 206 professionals named Yuming Zhou, who use LinkedIn to exchange ...

Yuming Zhou | Nanjing University, Nanjing | NJU | ResearchGate
https://www.researchgate.net/profile/Yuming_Zhou 翻译此页
Yuming Zhou of Nanjing University, Nanjing NJU. Contact Yuming Zhou on ResearchGate, the

2017年9月



2018年9月

文件(E) 编辑(E) 查看(V) 历史(S) 书签(B) 工具(T) 帮助(H)

Yuming Zhou_百度搜索

https://www.baidu.com/s?ie=utf-8&f=3&rsv_bp=0&rsv_idx=1&tn=baidu&wd=Yuming Zhou&rsv_pq=a82ae...

百度一下

百度首页 设置 登录

网页 资讯 知道 图片 视频 贴吧 文库 地图 音乐 更多»

百度为您找到相关结果约201,000个

[dblp: Yuming Zhou](#)

查看此网页的中文翻译, 请点击 翻译此页

List of computer science publications by Yuming Zhou... Zhifei Chen, Lin Chen, Wanwangyinq Ma, Xiaoyu Zhou, Yuming Zhou, Baowen Xu: Understanding metric-b...

[https://dblp.uni-trier.de/pers... - 百度快照](#)

[Yuming Zhou | 领英](#)

上领英, 在全球领先职业社交平台查看Yuming Zhou的职业档案。Yuming的职业档案列出了1个职位。查看Yuming的完整档案, 结识职场人脉和查看相似公司的职位。

[www.linkedin.com/in/ym... - 百度快照](#)

[Home page for Yuming Zhou](#)

查看此网页的中文翻译, 请点击 翻译此页

2011: Musheng Wang (MSc), Ming Yang (MSc, co-supervised with Prof. Xiaoyu Zhou) Selecte d Papers[1] Yuming Zhou, Yibiao Yang, Hongmin Lu, Lin Chen...

[cs.nju.edu.cn/zhouyuming - 百度快照](#)

[符合“Yuming Zhou”的 200+ 位会员 | 领英](#)

yuming zhou Student at hebeijingmao 所在地区 中国 河北 石家庄 教育经历 英国卡地夫大学

查看完整档案 Public Profile Yuming Zhou -- 所在地区 中国 查看完整档案 ...

[https://cn.linkedin.com/pub/d/... - 百度快照](#)

[zhouyuming_hbxt的博客 - 第 1 页 - CSDN博客](#)

CSDN博主 zhouyuming_hbxt的博客 主页提供丰富的内容介绍, 包含博客等级、博主粉丝、积 分、排名等内容, 查找最新 zhouyuming_hbxt的博客 博文更新信息, 请上CSDN博客...

[https://blog.csdn.net/zym32697... - 百度快照](#)

[Yuming Zhou's research works | Nanjing University, Nanjing \(...\)](#)

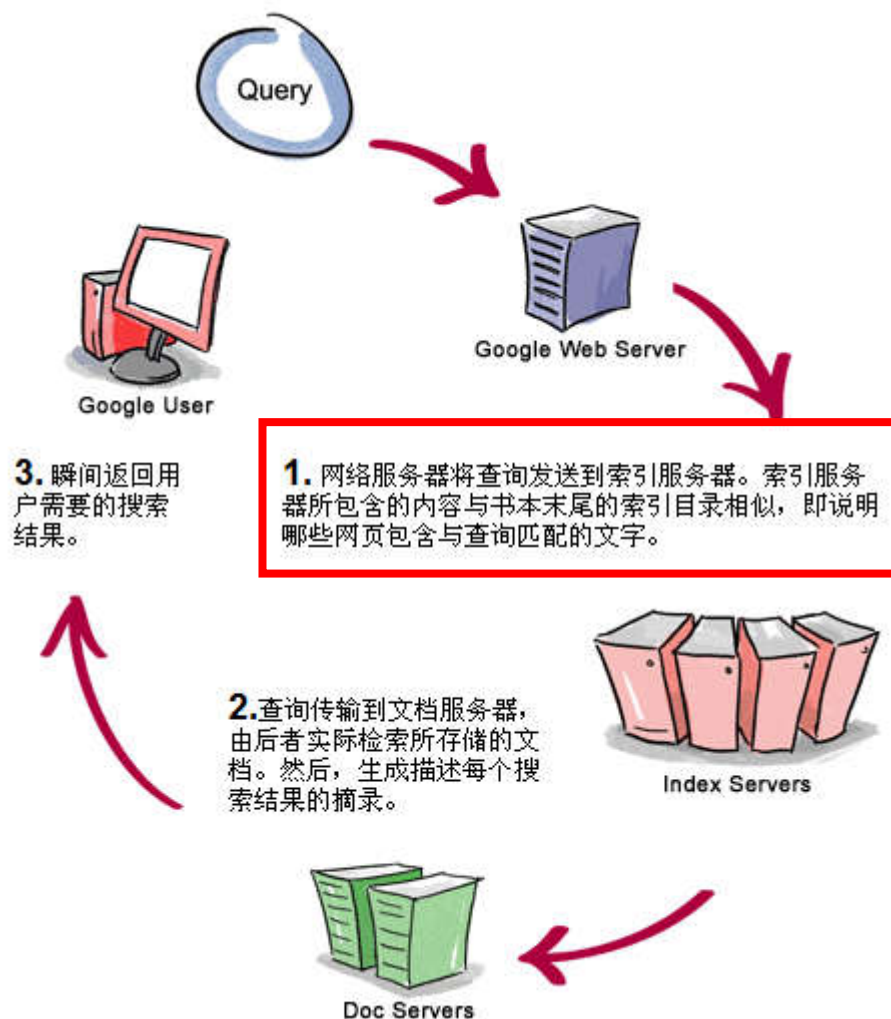
搜索热点

排名	搜索指数	
1	互联网法院 抖音	330209 ↑
2	马云新名片	271964 ↓
3	裸条贷女孩卖淫	253073 ↑
4	李小璐朋友圈曝光	251499 ↓
5	黄牛转战APP	206071 ↑
6	8万托关系上名校	203598 ↓
7	崔永元回应冯小刚	196401 ↑
8	女生众筹上大学	193928 ↑
9	iPhone双卡双待	191904 ↓
10	高晓松 丁香医生	189430 ↑

来源: 百度风云榜 - 实时热点

Google的“秘诀”

- Google 查询的全过程通常不超过半秒时间，但在这短短的时间内需要完成多个步骤，然后才能将搜索结果交付给搜索信息的用户。



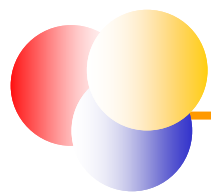
A

a-Platonic(ity), 125, 182, 283n, 284, 309
 academic libertarianism, 183, 307
 Aczel, Amir, 128
 aesthetics, 62, 99, 138, 253–73, 296, 297
 Al-Ghazali, 47, 171n
 Alpher, Ralph, 168
 Amioun, 4, 45, 65, 140, 154, 155, 171, 258, 287, 288
 Anderson, Chris, 223
 Apelles-style strategy, 204, 295, 307
 Apelles the Painter, 204, 295
 Aristotle, 202
 Aron, Raymond, 12
 Arrow, Kenneth, 283
 autism, 194, 323
 Averroës, 47, 171n

B

Bachelier, Louis, 282
 Bacon, Sir Francis, 101, 102, 167
 Ball, Philip, 267, 270
 Balzac, Honoré de, 103, 104, 105
 banks, 43, 123, 208, 225–26

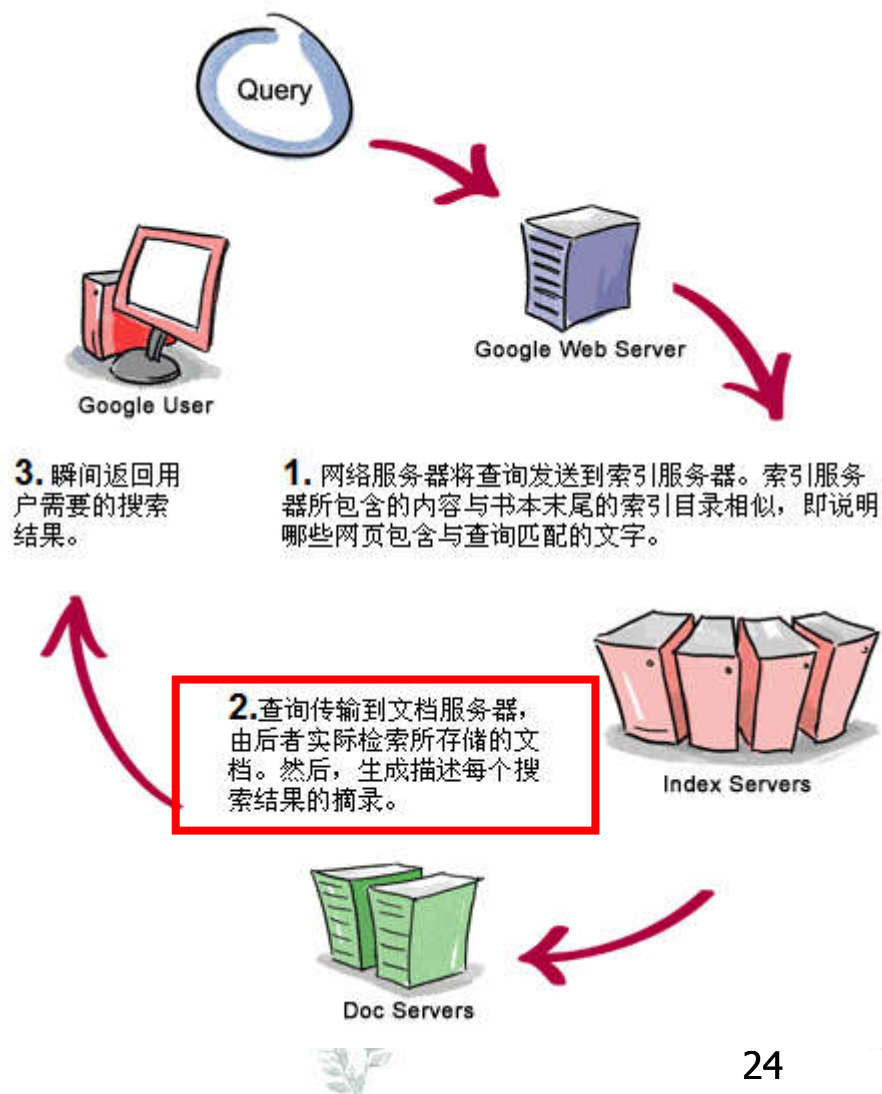
Barabasi, Albert-Laszlo, 226
 barbell strategy, 205–6, 207, 307
 Barron, Gerg, 77
 Bastiat, Frédéric, 111, 112, 288
 Bateson, Gregory, 25
 Baudelaire, Charles, 71
 Baumol, William, 90
 Bayle, Pierre, 48, 296
 bell curve, *see* Gaussian distribution
 Bernard, Claude, 278
 Berra, Yogi, 136, 199n, 208
 Berry, Michael, 178
 Bethe, Hans, 168
 bildungsphilister, 131, 255, 295, 307, 319
 bin Laden, Osama, 16
 biology, 219, 328
 Black Swan, defined, xvii–xviii, xix, xx–xxi, xxv, xxvii
 Black Swan blindness, xxiii, 77–79, 141–42, 288, 307
 Black Swan ethical problem, defined, 308
 Bloch, Marc, 101
 Bohr, Niels, 136n



Google的“秘诀”

- Google 查询的全过程通常不超过半秒时间，但在这短短的时间内需要完成多个步骤，然后才能将搜索结果交付给搜索信息的用户。

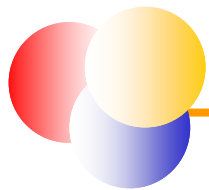
PageRank?



Google的“秘诀”

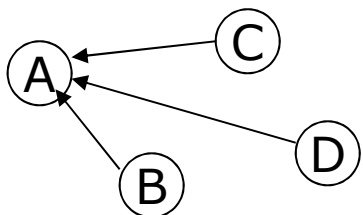
- 它利用网页的link结构计算网页的rank，用来量化网页的“重要性”，称为PageRank
 - PageRank是Google的trademark
- Google利用link改进网页的搜索引擎



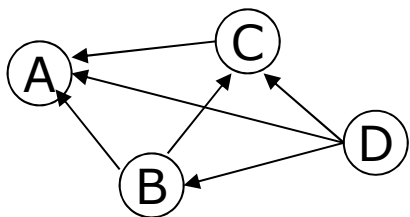


Google的“秘诀”

- 假定4个网页：**A**, **B**, **C** 和**D**.



$$PR(A) = PR(B) + PR(C) + PR(D).$$

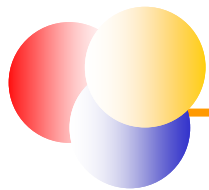


$$PR(A) = \frac{PR(B)}{2} + \frac{PR(C)}{1} + \frac{PR(D)}{3}.$$

- 假定 $L(A)$ 是网页A的out link数目，那么

$$PR(A) = \frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)}.$$

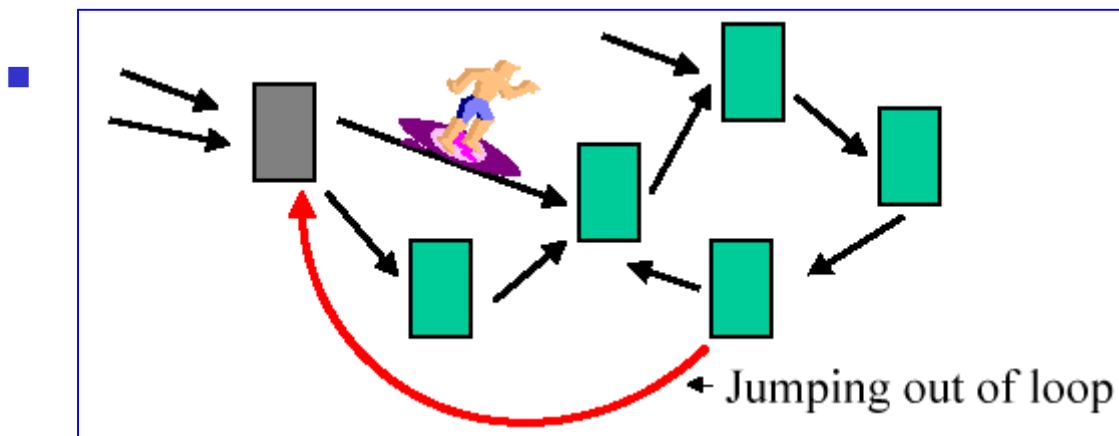




Google的“秘诀”

- 假定A有in link B, C, D ...。一般情况下，阻尼系数 $d = 0.85$

$$PR(A) = 1 - d + d \left(\frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)} + \dots \right)$$



。访问一个网页时，
页的访问，最终因厌

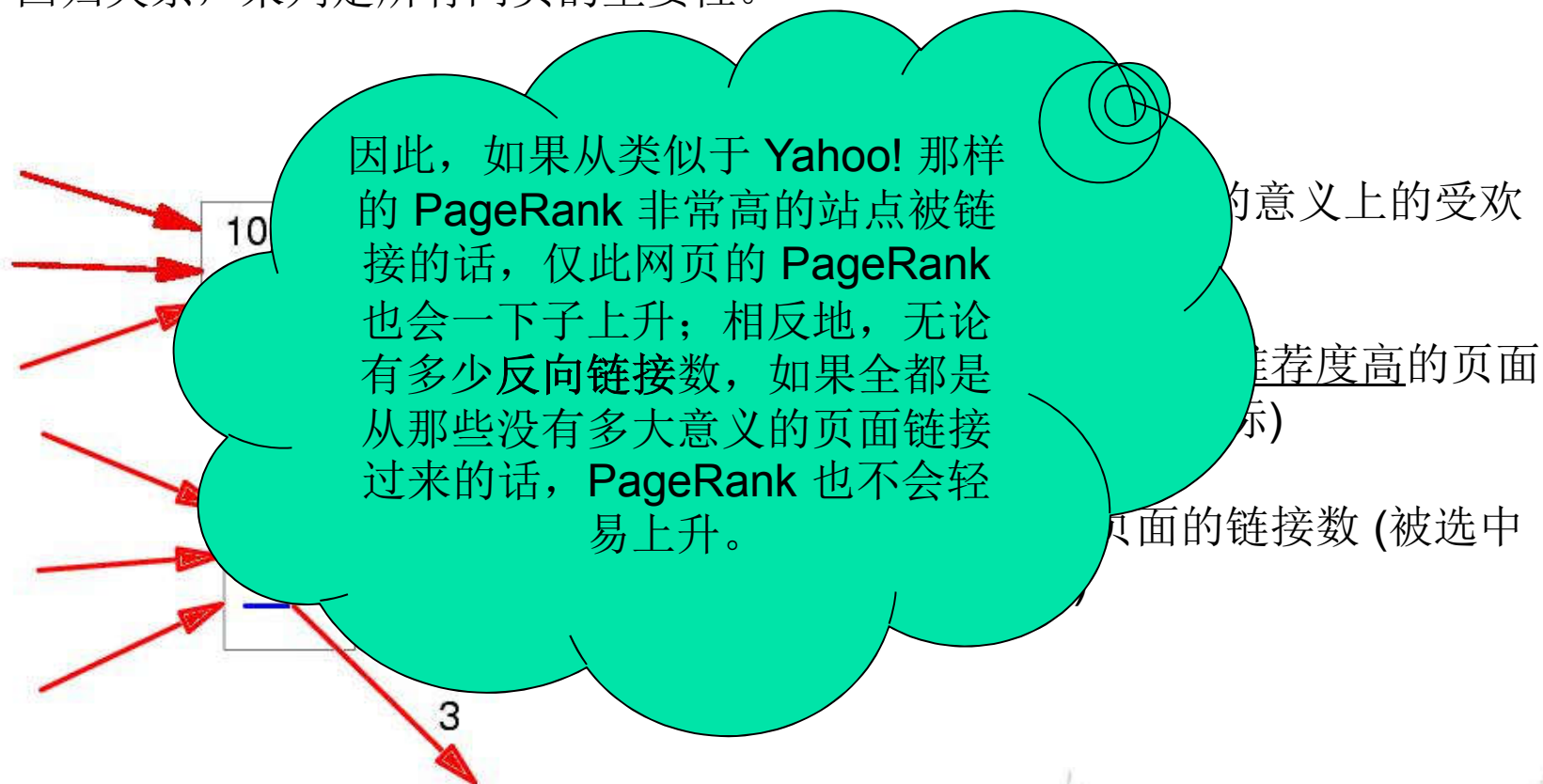
它的PageRank.

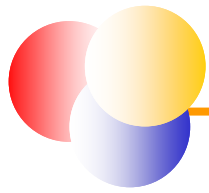
网页的概率

- 一个网页具有高的PageRank，如果
 - 有许多网页指向它
 - 或者只有一些网页指向它，但这些网页都具有高的PageRank

Google的“秘诀”

- PageRank 是基于「从许多优质的网页链接过来的网页，必定还是优质网页」的回归关系，来判定所有网页的重要性。





Google的“秘诀”

$$R_0 \leftarrow S$$

- initialize vector over web pages

loop:

$$R_{i+1} \leftarrow A^T R_i$$

- new ranks sum of normalized backlink ranks

$$d \leftarrow \|R_i\|_1 - \|R_{i+1}\|_1$$

- compute normalizing factor

$$R_{i+1} \leftarrow R_{i+1} + dE$$

- add escape term

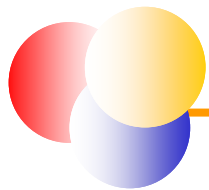
$$\sigma \leftarrow \|R_{i+1} - R_i\|$$

- control parameter

while $\sigma > \varepsilon$

- stop when converged

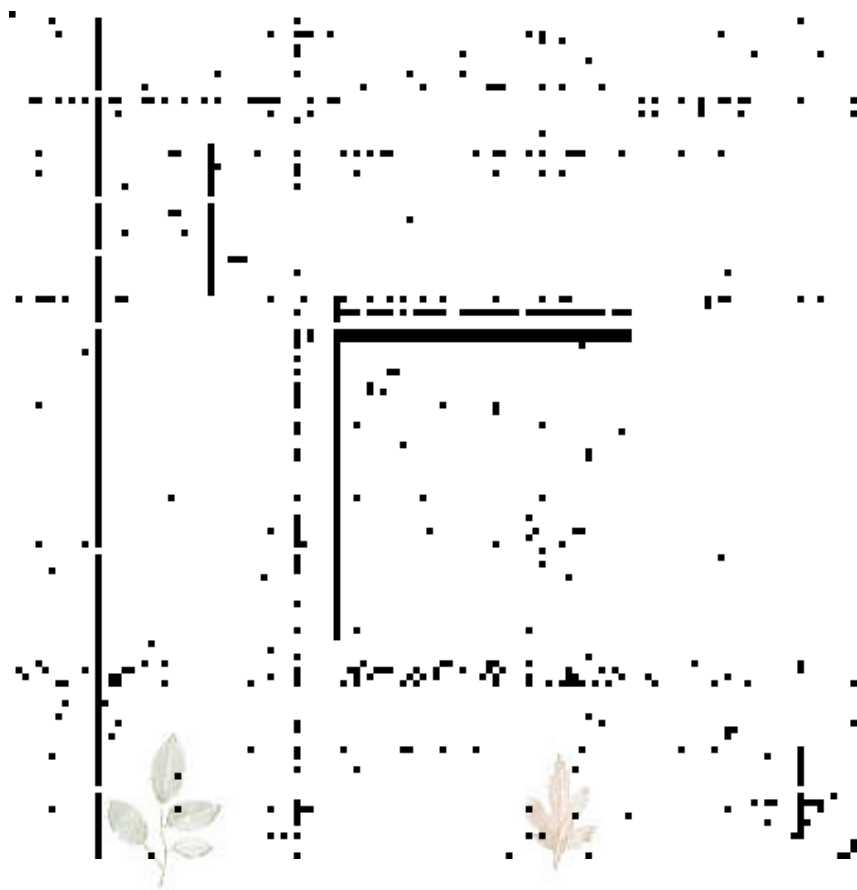




Google的“秘诀”

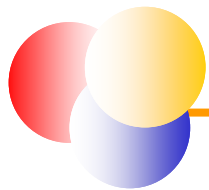
行列阵的形式

$a_{ij}=1$ if (从页面 i 向页面 j 「有」 链接的情况)
 $a_{ij}=0$ if (从页面 i 向页面 j 「没有」 链接的情况)



- 当黑点呈横向排列时，表示这个页面有很多正向链接(即向外导出的链接);
- 反之，当黑点呈纵向排列时，表示这个页面有很多反向链接。

PageRank 的行列阵是把这个邻接行列倒置后 (行和列互换)，为了将各列(column)矢量的总和变成 1 (全概率)，



Google的“秘诀”

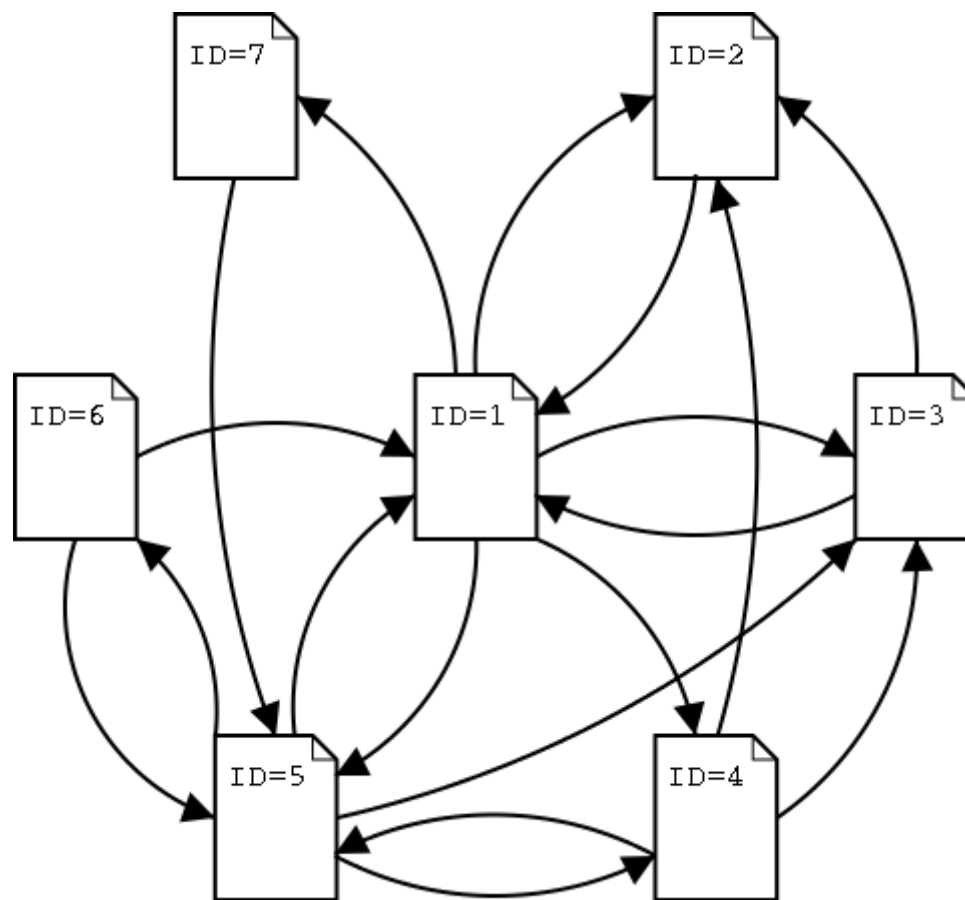
PageRank实例

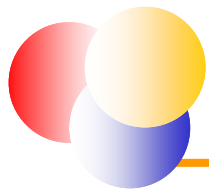
链接源ID

1
2
3
4
5
6
7

链接目标ID

2,3,4,5,7
1
1,2
2,3,5
1,3,4,6
1,5
5

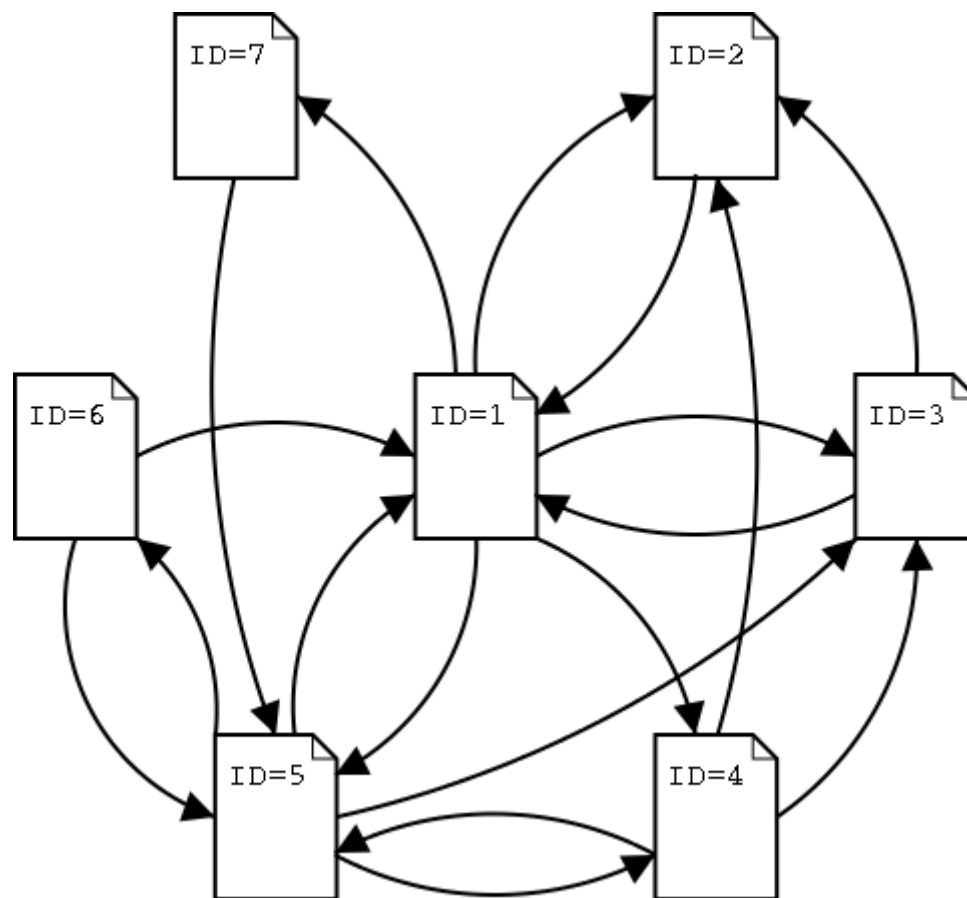


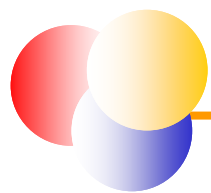


Google的“秘诀”

PageRank实例

$A = \begin{bmatrix} & 1, 2, 3, 4, 5, 6, 7 \\ 1, & 0, 1, 1, 1, 1, 0, 1; \\ 2, & 1, 0, 0, 0, 0, 0, 0; \\ 3, & 1, 1, 0, 0, 0, 0, 0; \\ 4, & 0, 1, 1, 0, 1, 0, 0; \\ 5, & 1, 0, 1, 1, 0, 1, 0; \\ 6, & 1, 0, 0, 0, 1, 0, 0; \\ 7, & 0, 0, 0, 0, 1, 0, 0; \end{bmatrix}$





Google的“秘诀”

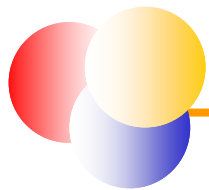
PageRank实例

- M: 将 A 倒置后将各个数值除以各自的非零要素

M = [

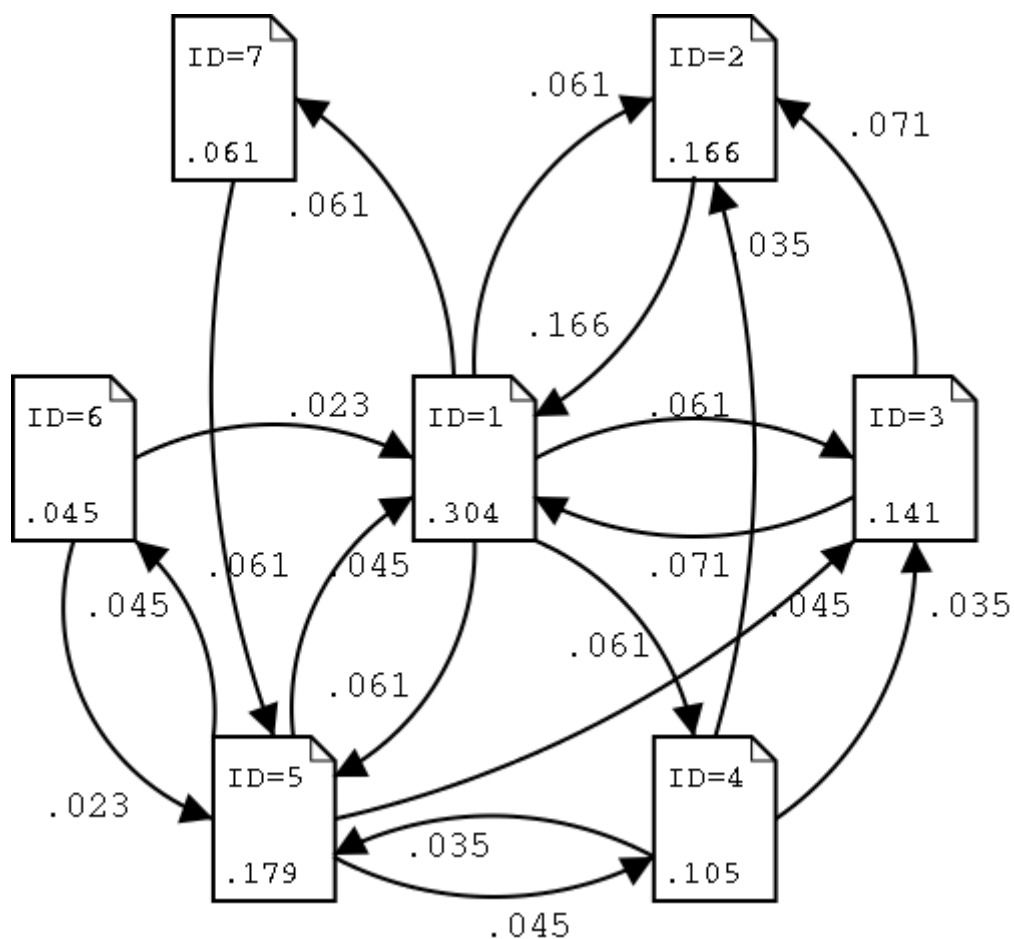
0,	1,	1/2,	0,	1/4,	1/2,	0;
1/5,	0,	1/2,	1/3,	0,	0,	0;
1/5,	0,	0,	1/3,	1/4,	0,	0;
1/5,	0,	0,	0,	1/4,	0,	0;
1/5,	0,	0,	1/3,	0,	1/2,	1;
0,	0,	0,	0,	1/4,	0,	0;
1/5,	0,	0,	0,	0,	0,	0;

]



Google的“秘诀”

PageRank实例



- 流入量
=(ID=2发出的Rank)
+(ID=3发出的Rank)
+(ID=5发出的Rank)
+(ID=6发出的Rank)
=
 $0.166 + 0.141/2 + 0.179/4 + 0.045/2$
= 0.30375

第六章 排序 (PageRank)

王海

<http://bigdata.lic.nkfust.edu.tw/ezfiles/141/1141/img/1900/227221861.ppt>



PageRank算法

- ➡ 基本PageRank
- ➡ 面向主题PageRank
- ➡ Link Spam与反作弊
- ➡ 导航页与权威页

一小组：王高翔，李渠，刘晴，柳永康，刘昊骋
二小组：王飞，李天照，赵俊杰，陈超，陈瑾翊



<https://wenku.baidu.com/view/77de601de2bd960590c67767.html>



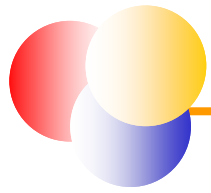


Google的“秘诀”

- 如何证明PageRank的“迭代”计算收敛？
- 如何高效地计算PageRank(时间和空间)，特别是网页数量巨大的时候？
- 如何设计网站，使得其网页排名靠前？

- ① M. Bianchini et al. [Inside PageRank](#). ACM Transactions on Internet Technology, 2005.
- ② A. Langville, C. Meyer. [Deeper Inside PageRank](#). Internet Mathematics, 2003.





PageRank的应用：关键模块识别

单元测试中的一个问题

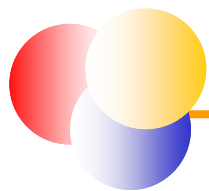
- (1) 所测试的系统规模大，模块数目很多
- (2) 时间非常紧，人手不够用

不可能测试所有模块，那么应该优先测试哪些模块？

我们的建议：

- (1) 优先测试“关键”模块
- (2) 优先测试“最有可能包含bug”的模块

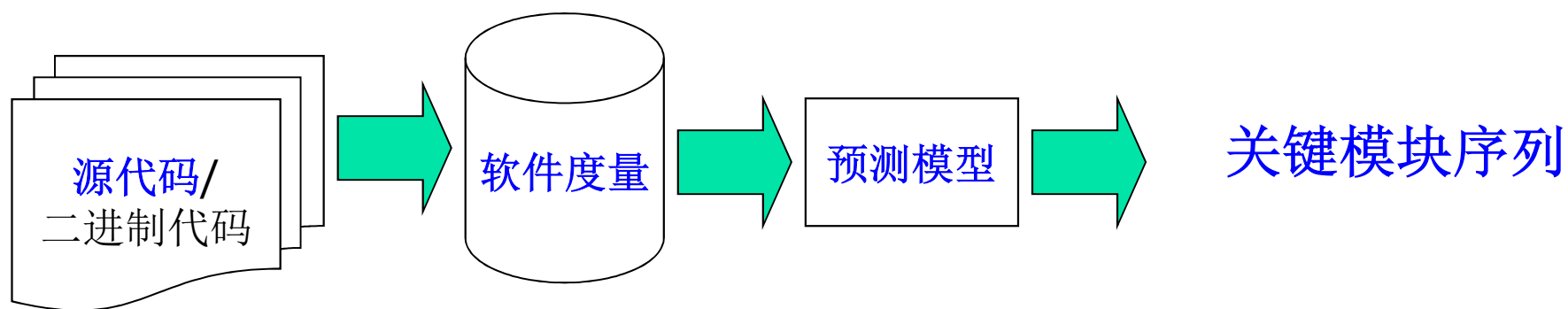


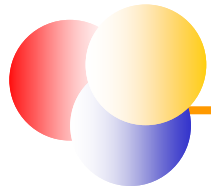


PageRank的应用:关键模块识别

如何识别系统中的“关键”模块？

我们的基本思路：





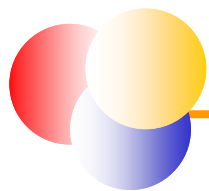
PageRank的应用:关键模块识别

什么是关键类?

“The classes that implement the key concepts of a system. ... they manage a large number other classes or use them in order to implement their function...” ----Tahvildari et al. JSME 2004

关键类不是 “God classes”





PageRank的应用:关键模块识别

如何识别关键类?

- 方法1: 阅读文档
文档有可能已经过时

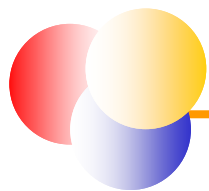
- 方法2: 手动审查所有的类

对大的系统而言, 在时间紧张的情况下,
不太可能审查系统中所有的类

我们的目标:

对给定源代码的系统, 自动识别出关键类





PageRank的应用:关键模块识别

Step 1:

源代码

分析

类间依赖图

Step 2:

计算

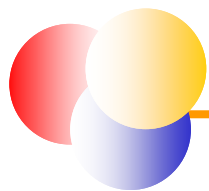
Web度量

Step 3:

排序

候选的关键类





PageRank的应用:关键模块识别

CDG (Class dependence graph)

$$G = (N, E, W)$$

- 节点集N:

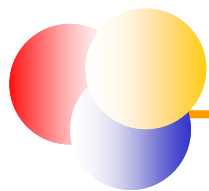
- 类
- 接口

- 边集E:

- Extension relationship
- Implementation relationship
- Types of attributes

- 权集W: 依赖关系的数目

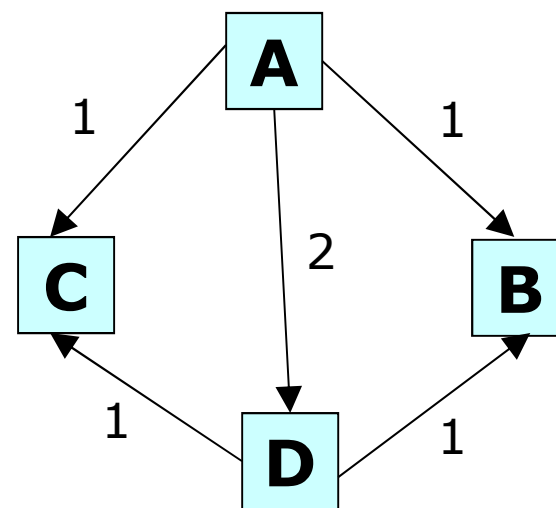


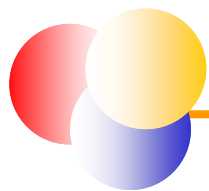


PageRank的应用:关键模块识别

```
public class A extends B implements C{  
    private D attribute1;  
    private D attribute2;  
    ...  
}  
public class B {  
    ....  
}  
public class D implements C{  
    private B attribute1;  
    ...  
}  
public interface C{  
    ...  
}
```

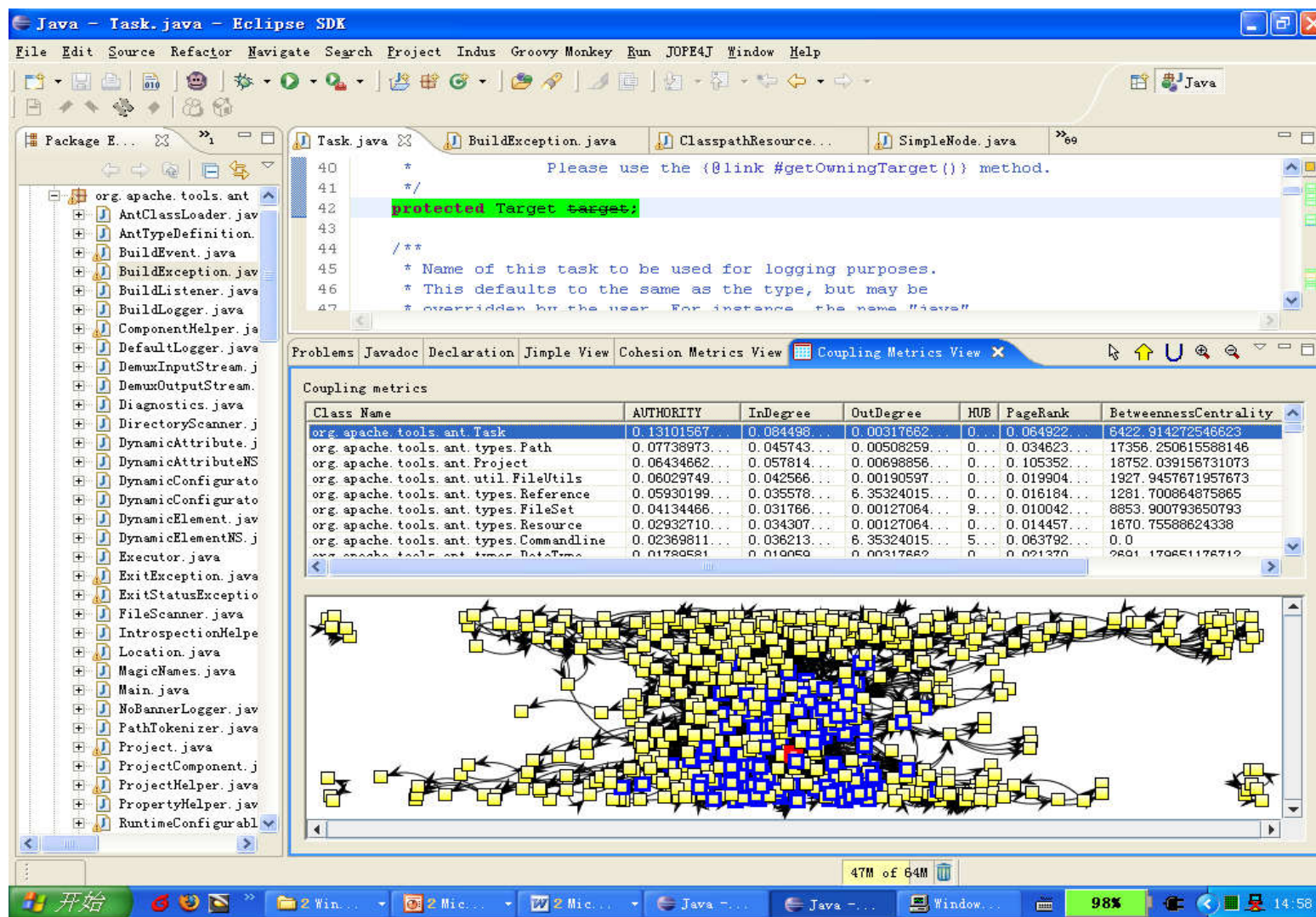
CDG举例

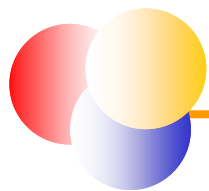




PageRank的应用:关键模块识别

Ant 1.6.1





PageRank的应用:关键模块识别

CDG: 形式化的定义

对一个系统 s , $G = (N, E, W)$

$$N = \{c \in Classes(s) \mid \exists c' (c' \in Classes(s) \wedge c \neq c' \wedge CI(c, c') + CA(c, c') > 0)\}$$

$$E = \{ \langle c_1, c_2 \rangle \mid c_1, c_2 \in N \wedge CI(c_1, c_2) + CA(c_1, c_2) > 0 \}$$

$$W : E \rightarrow R^+$$

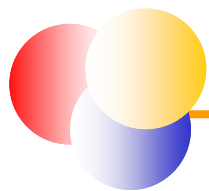
$$W(\langle c_1, c_2 \rangle) = CI(c_1, c_2) + CA(c_1, c_2)$$

Where

$CI(c, c')$: the number of extension/implementation relationships from c to c'

$CA(c, c')$: the number of attributes in c whose type is c'





PageRank的应用:关键模块识别

(1) Pagerank: 加权的版本

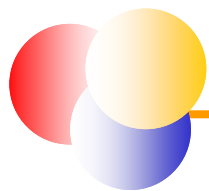
$$r(c) = d * \sum_{c' \in pred(c)} \frac{W(<c', c>) * r(c')}{|succ(c')|} + (1 - d)$$

(2) HITS: 加权的版本

$$a(c) = \sum_{c' \in pred(c)} W(<c', c>) * h(c') \quad \textbf{Authority}$$

$$h(c) = \sum_{c' \in succ(c)} W(<c, c'>) * a(c') \quad \textbf{Hubiness}$$





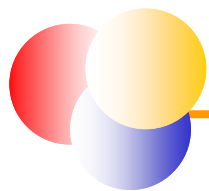
PageRank的应用:关键模块识别

实验分析

系统	类的数目	总代码行数	真实的关 键类
Ant 1.6.1	664	76K	10
JMeter 2.0.1	706	73K	14

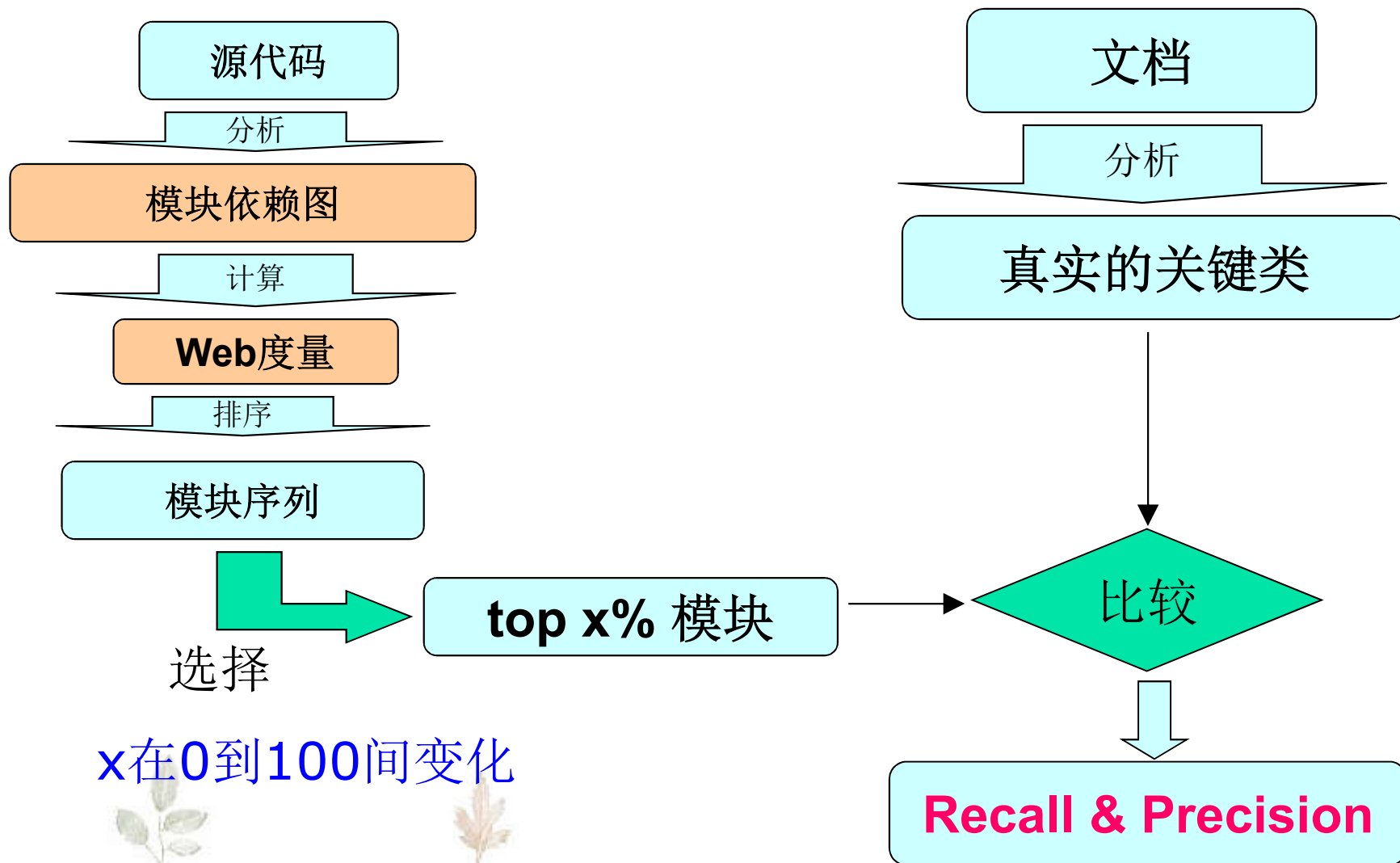
**Identified from documents by
Zaidman et al. (JSME 2008)**

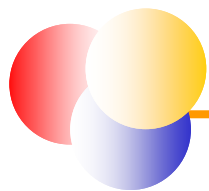




PageRank的应用:关键模块识别

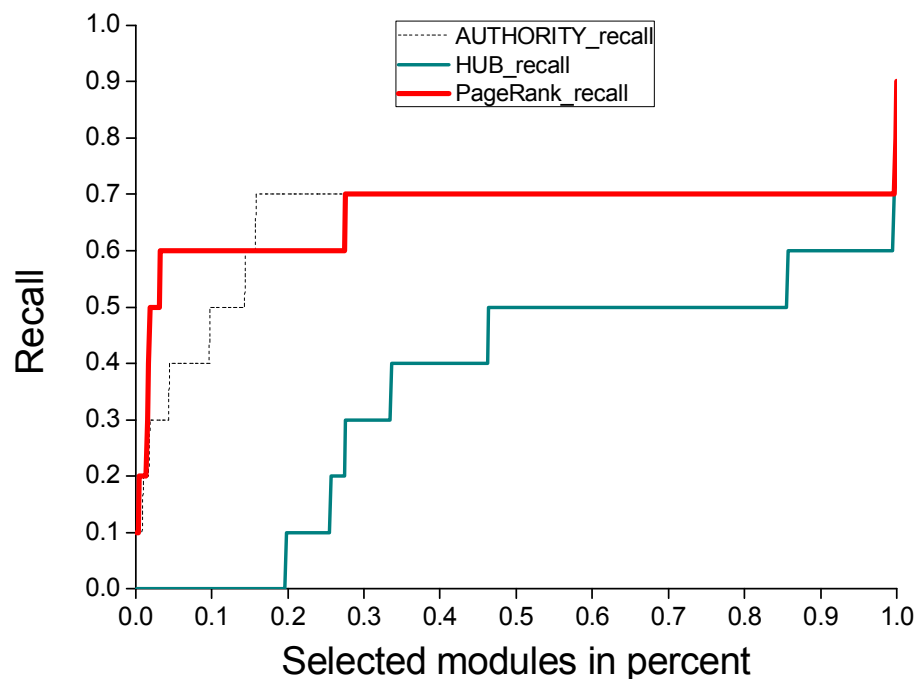
研究方法



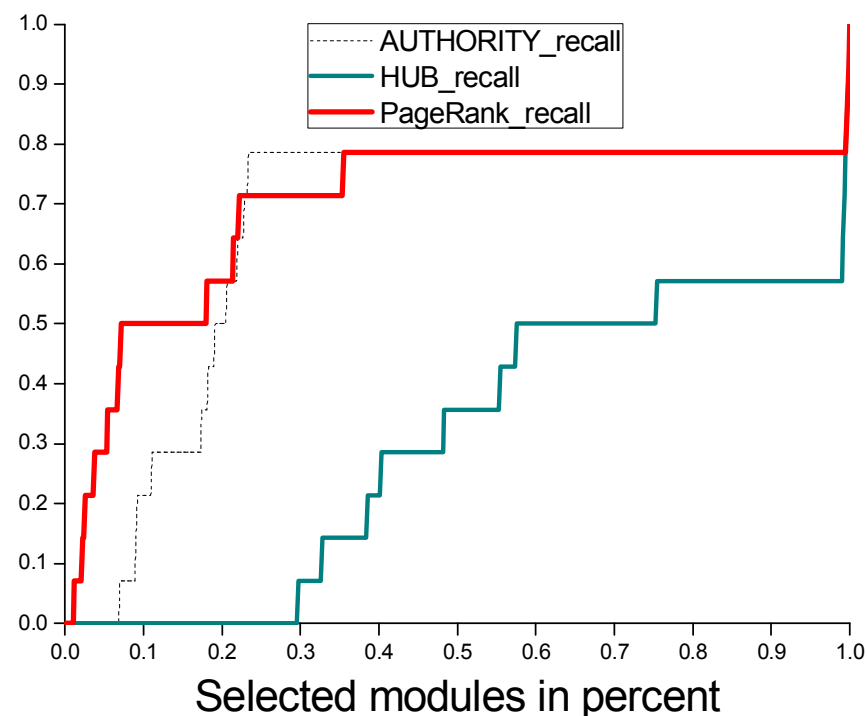


PageRank的应用:关键模块识别

Recall

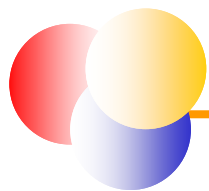


Ant 1.6.1



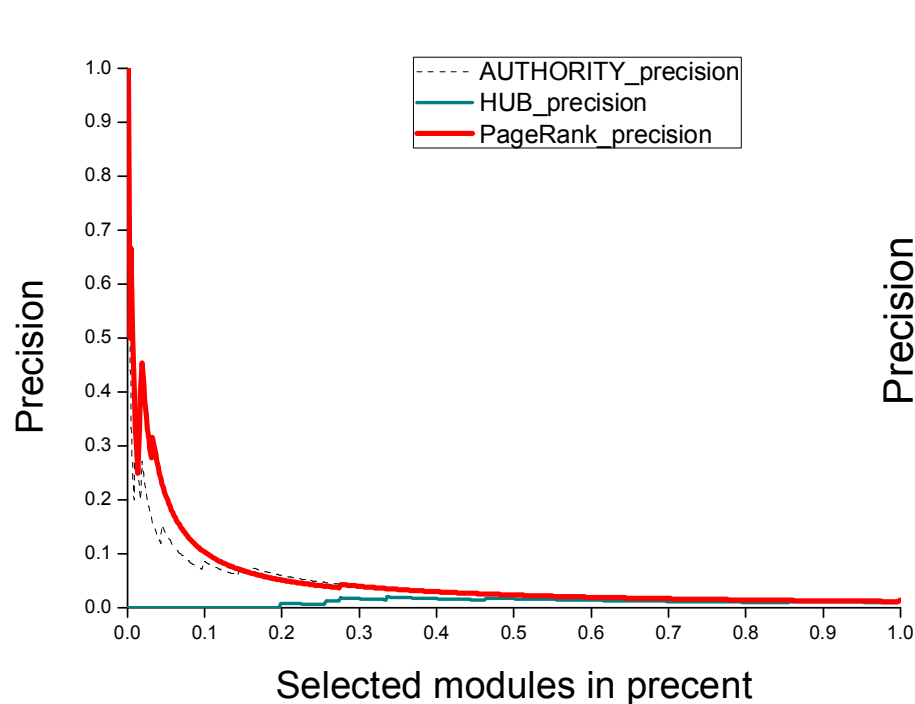
Jmeter 2.0.1



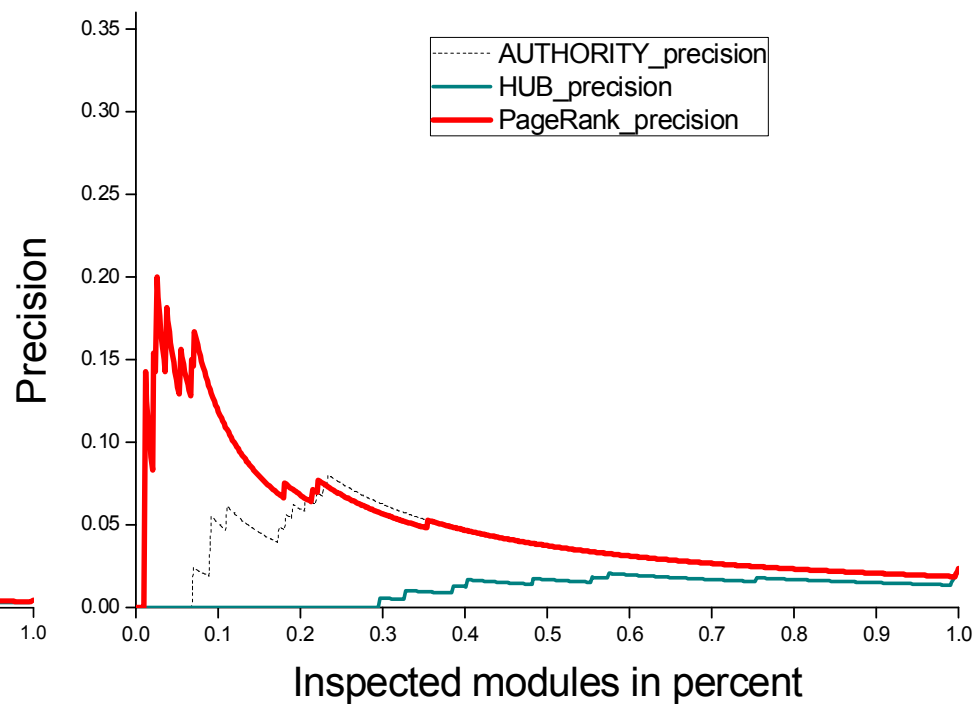


PageRank的应用:关键模块识别

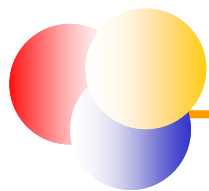
Precision



Ant 1.6.1

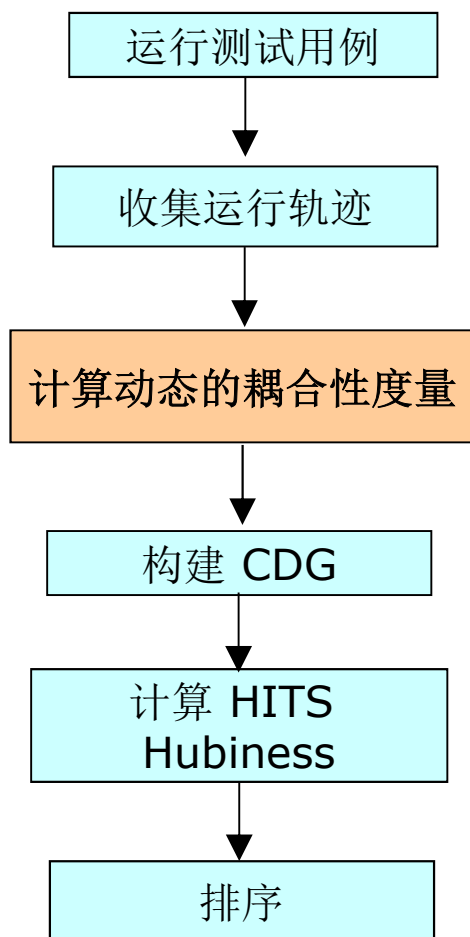


Jmeter 2.0.1



PageRank的应用:关键模块识别

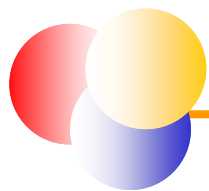
Zaidman等的“动态”方法



选择top 15%的类进行审查

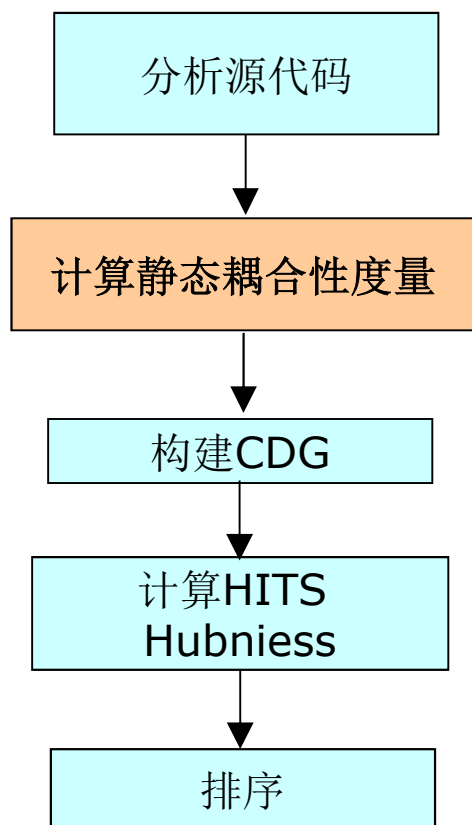
System	Recall	Precision	Trace	Time
Ant 1.6.1	0.9	0.47	2GB	1h 45min
JMeter 2.0.1	0.93	0.46	600MB	1h 15min

- 需要好的测试集
- 需要收集运行轨迹
- 需要大的存储空间
- 需要很长时间



PageRank的应用:关键模块识别

Zaidman等的“静态”方法

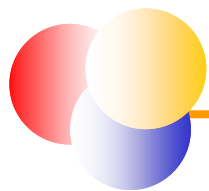


选择top 15%的类进行审查

System	Recall	Precision	Time
Ant 1.6.1	0.5	0.08	1h 01min
JMeter 2.0.1	0.43	0.08	1h 31min

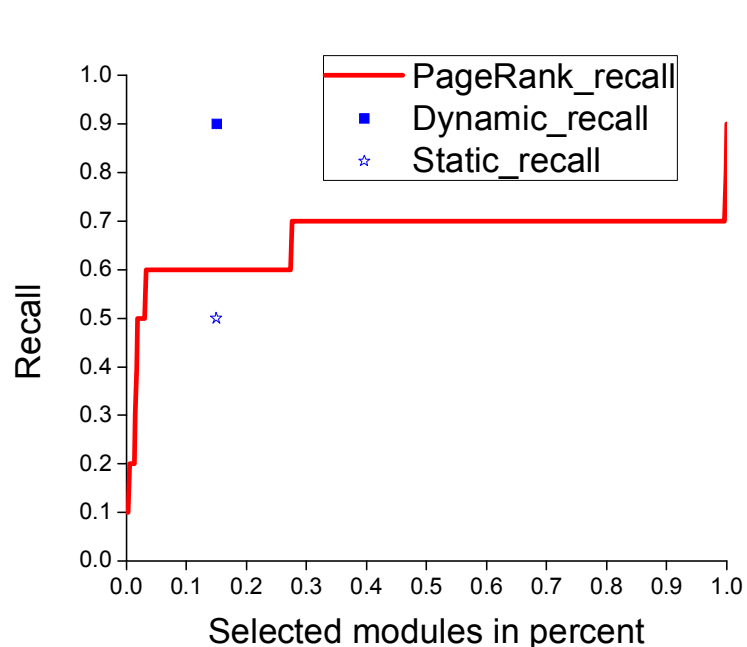
- 低的recall和precision
- 需要很长时间

Resemble dynamic metrics by counting the number of polymorphic calls between classes

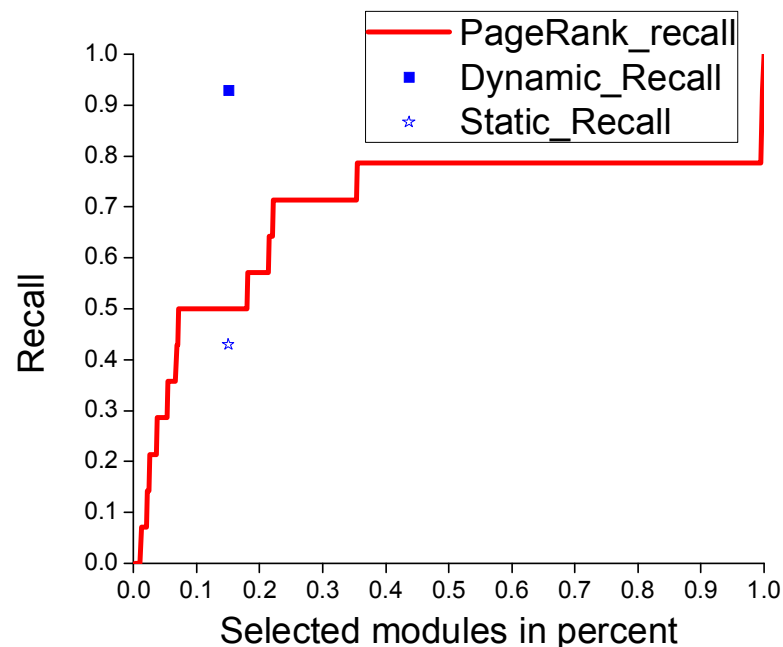


PageRank的应用:关键模块识别

相关工作比较: recall

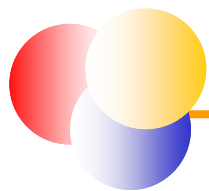


Ant 1.6.1



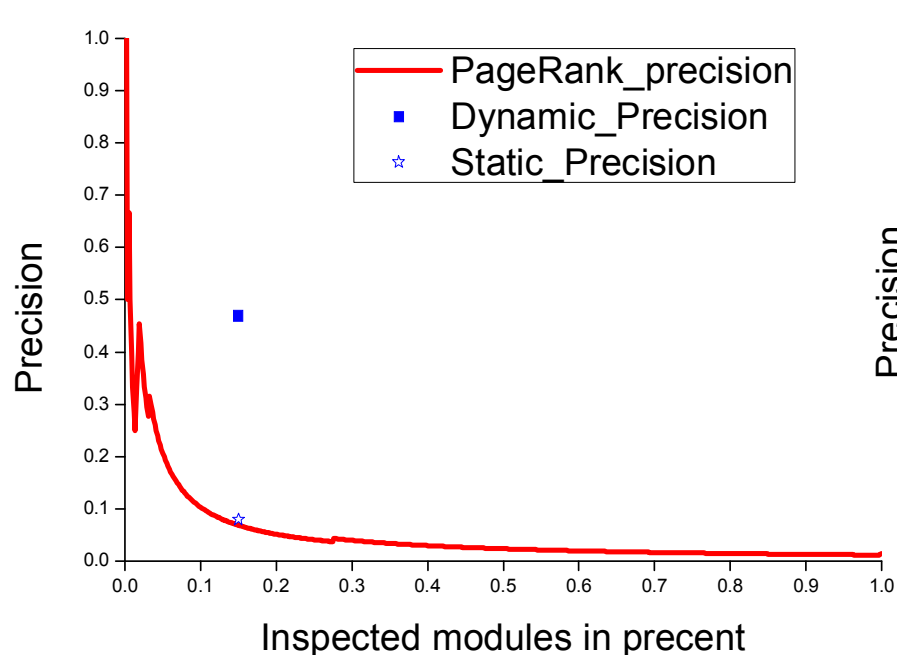
Jmeter 2.0.1



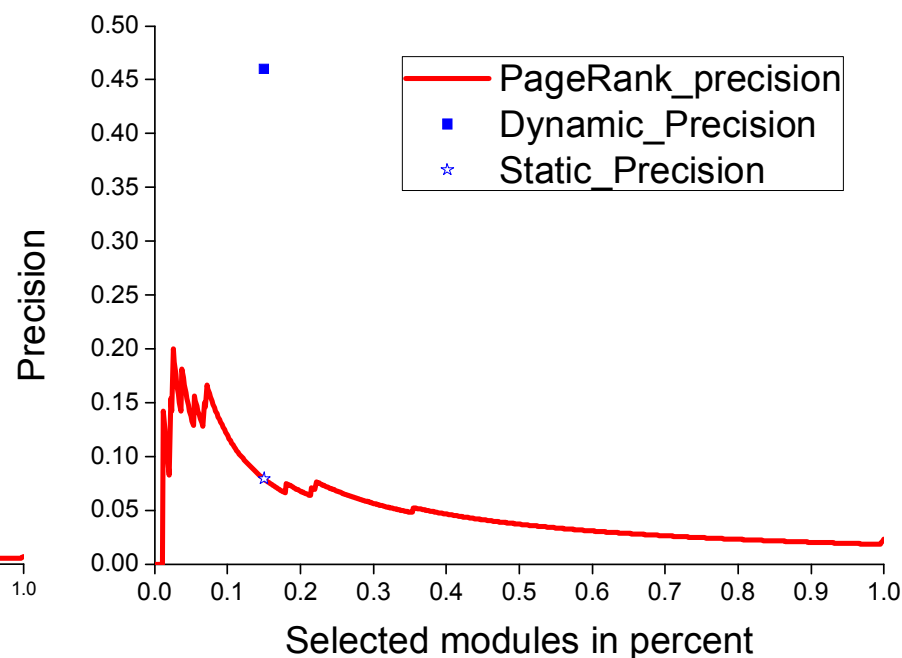


PageRank的应用:关键模块识别

相关工作比较: precision

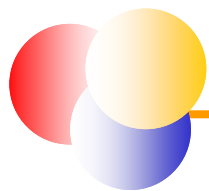


Ant 1.6.1



Jmeter 2.0.1





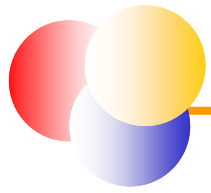
PageRank的应用:关键模块识别

相关工作比较: **time**

Zaidman等的方法 > 1 小时

我们的方法 < 10 秒

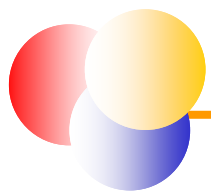




小结

- 量纲分析
- Google的“秘诀”





思考题

1. 如何防止淘宝上的“信誉欺诈”？
2. 如何确定软件系统的最佳发布日期？



**Thanks for your time and
attention!**

