

DOI:10.11992.tis.201603045
网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.TP.20160513.0923.022.html>

基于稠密子图的社区发现算法

郑文萍^{1,2}, 张浩杰¹, 王杰^{1,2}

(1.山西大学 计算机与信息技术学院,山西 太原 030006; 2.山西大学 计算智能与中文信息处理教育部重点实验室,山西 太原 030006)

摘 要:基于密度的图聚类算法在社区发现中得到了广泛应用,然而由于其通过搜索网络中局部稠密子图来识别社区,使得大量结点因不能构成稠密子图而未被聚类。针对此问题,给出了一种基于稠密子图的软聚类算法 (community detection based dense subgraphs, BDSG)。首先给出一种中心社区发现方法;进而定义了一种结点的社区归属感,并给出中心社区扩展策略;最终得到聚类结果。通过与 CPM (clique percolation method)、 k -dense 算法在空手道俱乐部、海豚社交网络、大学生足球网络、电子邮件网络和合作网络等数据进行比较,表明 BDSG 算法在模块性指标与时间效率方面体现了良好性能,同时中心社区扩展策略能在一定程度上提高 CPM、 k -dense 等基于密度算法的聚类有效性。

关键词:复杂网络;社区发现;图聚类;软聚类;密度;中心扩展策略;点介数;模块性
中图分类号:TP18 **文献标志码:**A **文章编号:**1673-4785(2016)03-0426-07.

中文引用格式:郑文萍,张浩杰,王杰.基于稠密子图的社区发现算法[J]. 智能系统学报, 2016, 11(3): 426-432.
英文引用格式:ZHENG Wenping, ZHANG Haojie, WANG Jie. Community detection algorithm based on dense subgraphs[J]. CAAI transactions on intelligent systems, 2016, 11(3): 426-432.

Community detection algorithm based on dense subgraphs

ZHENG Wenping^{1,2}, ZHANG Haojie¹, WANG Jie^{1,2}

(1. School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China; 2. Key Laboratory of Computation Intelligence and Chinese Information Processing, Ministry of Education, Shanxi University, Taiyuan 030006, China)

Abstract:The density-based graph clustering algorithm has been widely used in community detection. However, because it identifies a community by searching a partially dense subgraph in the network, many nodes do not constitute a dense subgraph and are therefore difficult to cluster. In this paper, we present a soft clustering algorithm based on dense subgraphs (BDSG) for detecting communities in complex networks. First, we propose a method for detecting the central communities. Next, we define the degree of community attribution of a node, and put forward a core community extended strategy. Finally, we obtain the clustering results of a network. Compared with the clique percolation method (CPM), k -dense algorithms from Zachary's Karate Club, the dolphin social network, the American college football network, the email network, and the collaboration network, BDSG shows considerably better performance with respect to modularity and time efficiency. In addition, the proposed core community extended strategy may improve the effectiveness of the clustering-methods-based density, such as that in CPM, k -dense algorithms, and others.

Keywords:complex network; community detection; graph clustering; soft clustering; density; core extended strategy; vertex betweenness; modularity

象。大量研究表明,复杂网络中存在一种普遍特征——社区结构^[4]。复杂网络中社区发现^[5]不仅有助于深入研究整个网络的拓扑结构、功能模块以及动力学特性,同时在生物蛋白质的性能与互作用的分析^[6]、社会组织结构的网络分析^[7]、搜索引擎^[8]及推荐系统^[9]等方面均有广泛的应用前景,因此具有十分重要的理论意义和应用价值。

目前,社区发现算法的研究主要分为基于图划分的聚类算法^[10-11]、基于谱分析的聚类算法^[12]、基于层次的聚类算法^[13]和基于密度的聚类算法^[14]等。其中基于密度的聚类算法通过搜索网络中稠密子图^[15]能较好地发现网络中的功能模块,因此在社区发现中得到了广泛应用。2005年,Palla等^[16]提出派系过滤算法(clique percolation method, CPM),首先挖掘网络中结点数大于 k 的所有派系(完全图),然后将重叠结点大于 $k-1$ 的派系合并得到 k 派系社区。2006年,Saito等^[17]提出 k -dense子图结构,通过寻找网络中的 k -dense结构进行社区检测。2009年,Sun等^[18]以CPM为基础,通过改进寻找派系的方法提高算法效率,提出迭代派系过滤算法(iterative-clique percolation method, ICPM)。2010年,Liu等^[19]提出基于极大团的聚类算法(clustering-based on maximal cliques, CMC),通过搜索网络中的所有极大团,并依据相互连接度合并重叠率较高的极大团得到网络的社区结构。由于这些算法要搜索网络中的相对稠密子图来进行聚类,当网络中存在包含大量结点的稀疏子图时,这些结点可能最终成为未聚类结点,造成了聚类结果的不完全覆盖。这些未聚类结点构成的稀疏子图可能具有某种功能,或者与某些稠密子图共同行使功能,因此需要对网络中的部分未聚类结点进行进一步分析,判断其是否属于某一社区或形成新的社区。

针对基于密度算法中大量未聚类结点问题,提出一种基于稠密子图的社区发现算法(community detection based on dense subgraphs, BDSG)。首先通过搜索网络中的相对稠密子图得到中心社区;对于未聚类结点,定义了结点 v 对社区 C 的归属度 $b(v, C)$ 来度量结点和社区的连接倾向程度;基于归属度,给出一种中心社区扩展策略(core community extended strategy, CE),对未聚类结点进一步处理。BDSG算法中,一个结点可能属于多个社区,是一种软聚类方法。通过在空手道俱乐部、海豚社交网络、大学生足球网络、电子邮件网络和合作网络5个真实网络上与CPM、 k -dense算法进行比较,评估和分析BDSG算法在未聚类结点分配和社区模块性等方

面的性能表现。基于归属度的中心社区扩展策略也将应用在CPM、 k -dense等基于密度的图聚类算法中,对未聚类结点进一步处理,以提高聚类有效性。

1 背景知识

通常,一个复杂网络可以表示为图 $G=(V, E)$,其中顶点集 $V=\{v_1, v_2, \dots, v_n\}$, $n=|V|$;边集 E 中每条边 $e_{i,j}$ 对应 V 中一对顶点 (v_i, v_j) 之间的连接关系, $m=|E|$ 。顶点 v 的邻域 $N_G(v)=\{u \mid (v, u) \in E\}$,表示图 G 中与顶点 v 相邻的顶点集合,简记为 N_v 。结点 v 的度记为 k_v 。除非特别指明,以下仅考虑简单无向图,因此 $k_v=|N_v|$ 。令 $U \subseteq V(G)$,用 $[U]G$ 表示 G 的结点子集 U 的导出子图,在不发生混淆时,记为 $[U]$ 。记顶点子集 $[U]$ 在 G 中的邻域为 $N_G(U)=\{u \mid u \in N_G(x) \wedge x \in U\}$ 。

在复杂网络中,图 G 的密度^[20]记为 $D_G = \frac{m}{n(n-1)/2}$ 。可以看出, $D_G \in [0, 1]$,当 D_G 越趋近于1,图 G 中的边数越多;当 $D_G=1$ 时,图 G 为完全图。

结点 v_k 的点介数^[21] $B(v_k)$ 可以用来度量结点 v_k 在网络 G 中的重要性。如果一对结点 (v_i, v_j) 间共有 L_{v_i, v_j} 条不同的最短路径,其中有 $L_{v_i, v_j}(v_k)$ 条经过结点 v_k ,那么结点 v_k 对结点对 (v_i, v_j) 的贡献为 $L_{v_i, v_j}(v_k)/L_{v_i, v_j}$ 。定义结点 v_k 的点介数 $B(v_k)$:

$$B(v_k) = \sum_{i=1}^n \sum_{j=i+1}^n \frac{L_{v_i, v_j}(v_k)}{L_{v_i, v_j}} \quad (1)$$

通常一个结点的点介数越大,则该结点对网络结构的影响越大。点介数是网络中结点重要性度量指标之一。

2 结点对社区的归属度定义

基于密度的图聚类算法中可能存在大量不属于任何已有社区的未聚类结点,为了将这些结点聚类到合适的社区,需要定义未聚类结点和社区的关联强度,称为结点 v 对于社区 C 的归属度 $b(v, C)$ 。归属度的定义对聚类结果的影响至关重要,结点 v 对于社区 C 的归属度越大,则结点 v 属于社区 C 的可能性越大。

Cui等^[22]基于结点 v 与社区 C 关联边数定义了结点 v 对于社区 C 的归属度 $b_p(v, C) = \frac{|N_v \cap C|}{k_v}$,其中 $N_v \cap C = \{u \mid (v, u) \in E, u \in C\}$ 表示结点 v 在社区

C 中相邻点的集合, k_v 是结点 v 的度。然而结点 v 与社区 C 的关联强度不仅与关联边数有关, 也和社区 C 中 v 的相邻点在 C 中的重要性关系密切。

如图 1 所示, 当前聚类结果得到两个社区 C_1 (▲) 和 C_2 (■), 其余结点为未聚类结点。考虑结点 v_9 , 可得 $b_p(v_9, C_1) = b_p(v_9, C_2)$, $N_{v_9} \cap C_1 = \{v_{33}, v_{34}\}$, 且 $N_{v_9} \cap C_2 = \{v_1, v_3\}$ 。此时, 社区 C_1 中与结点 v_9 相邻结点的点介数比

例为
$$\frac{\sum_{u \in \{v_{33}, v_{34}\}} B(u)}{\sum_{w \in C_1} B(w)} = 0.95$$
, 而社区 C_2 中

与结点 v_9 相邻结点的点介数比例为
$$\frac{\sum_{u \in \{v_1, v_3\}} B(u)}{\sum_{w \in C_2} B(w)} = 0.83$$
。由于结点 v_9 在 C_1 中的相

邻点在网络中重要性更高, 可以认为 v_9 更倾向归属于社区 C_1 。

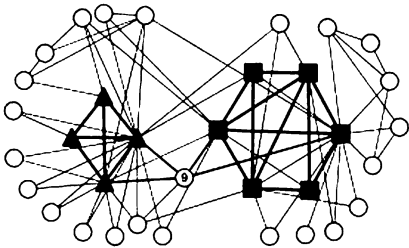


图 1 空手道俱乐部中未聚类结点分析
Fig.1 The analysis of subordinate vertices in zachary's karate club

因此, 度量未聚类结点和已有社区的归属度, 需要综合考虑该结点与一个社区关联边数以及社区内该结点的相邻结点的重要性。为了更准确地表示未聚类结点和社区的关系, 首先给出结点 v 对社区 C 的归属度定义:

$$b(v, C) = \alpha \times \frac{|N_v \cap C|}{k_v} + (1 - \alpha) \times \frac{\sum_{u_i \in N_v \cap C} B(u_i)}{\sum_{v_j \in C} B(v_j)}$$
 (2)

式(2)中第 1 项表示结点 v 与社区 C 关联边数, 第 2 项表示结点 v 连接社区 C 内结点的重要程度; $B(u_i)$ 表示结点 u_i 的点介数, 通过式(1)计算得到; $\alpha \in [0, 1]$ 为调节参数。 $b(v, C)$ 越大, 则 v 更倾向属于社区 C 。如果结点 v 在社区 C 中无相邻结点, 则 $b(v, C) = 0$ 。选择合理的调节参数 α 可以有效地减少最终聚类结果中的未聚类结点个数, 提高聚类效果, 表 1 给出了本文算法 BDSG 分别在 $\alpha = 0.8$ 和

$\alpha = 1$ (此时重要度对聚类结果不产生影响) 时的未聚类结点数 (subordinate vertices) 和模块性的比较结果, 表明通过重要度定义的归属度能够更加准确地表示节点和社区的关系。

表 1 不同 α 值时聚类结果的比较
Table 1 The comparison of the clustering results among different α

数据集	未聚类节点		Q	
	$\alpha=0.8$	$\alpha=1$	$\alpha=0.8$	$\alpha=1$
空手道俱乐部	1	3	0.820 5	0.717 9
海豚社交网络	0	1	0.773 5	0.761 0
大学生足球网络	0	0	0.639 0	0.615 0
电子邮件网络	34	41	0.722 4	0.715 1
合作网络	657	661	0.782 8	0.647 3

3 基于稠密子图的社区发现算法

基于稠密子图的社区发现算法 (BDSG) 主要由 2 部分构成: 首先通过搜索网络中大于指定密度阈值 d 的稠密子图得到网络中心社区, 确定聚类个数 k , 不属于任何一个中心社区的结点为未聚类结点; 根据式(2) 计算未聚类结点与已有社区的归属度, 将一些未聚类结点划分到归属度大于指定阈值的社区中, 对当前中心社区进行扩展; 更新剩余未聚类结点的归属度, 若网络中所有未聚类结点对任意社区的归属度都小于设定阈值, 则算法结束。

3.1 确定聚类个数

首先, 寻找网络中的子图密度大于指定阈值 d 的所有稠密子图。图 2 给出了 $d=0.9$ 时, 算法得到的 4 个稠密子图, 分别记为 U_1 、 U_2 、 U_3 和 U_4 。

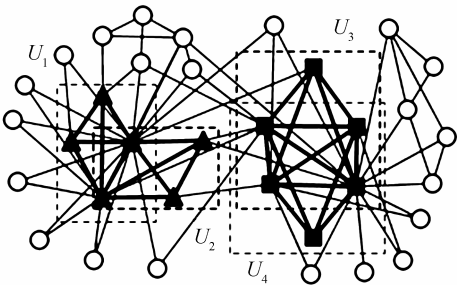


图 2 BDSG 算法在空手道俱乐部中得到的稠密子图
Fig.2 The dense subgraphs in zachary's karate club obtained by BDSG

然后, 建立子图重叠矩阵 M , 其中元素 $M_{i,i} =$

$|V(U_i)|$ 表示子图 U_i 中的顶点个数。元素 $M_{i,i} = |V(U_i) \cap V(U_j)|, i \neq j$, 表示子图 U_i 和 U_j 的公共顶点数。对 2 个不同稠密子图, 若 $M_{i,i} > \frac{\min(|U_i|, |U_j|)}{2}$, 则合并 2 个子图为一个新社区 $U' = [V(U_i) \cup V(U_j)]_c$, 此过程迭代进行, 直到不产生新的社区。得到的社区称为初始中心社区, 社区个数 k 为算法的聚类个数。图 2 中 4 个稠密子图的重叠矩阵表示为

$$M = \begin{bmatrix} 4 & 2 & 0 & 0 \\ 2 & 4 & 0 & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 4 & 5 \end{bmatrix}$$

进行稠密子图合并操作后可得到 2 个初始中心社区: $C_1 = [U_1 \cup U_2]_c, C_2 = [U_3 \cup U_4]_c$, 聚类个数 $k=2$ 。

算法确定了聚类个数和初始中心社区数之后, 不属于任何中心社区的结点就是未聚类结点。由于初始中心社区寻找过程中关注于网络中相对稠密的子图, 网络中存在大量未聚类结点, 需要设计合理的中心社区扩展策略, 对未聚类结点进一步处理。

3.2 中心社区扩展策略

设聚类个数为 k , 当前中心社区分别为 C_1, C_2, \dots, C_k , 则当前未聚类结点集合 $T = V(G) - \cup_{i=1}^k V(C_i)$ 。根据式(2)给出的结点对社区的归属度定义, 计算 T 中结点与中心社区 C_i 的归属度, 并对相关中心社区进行扩展, 具体过程如算法 1。

算法 1 中心社区扩展算法 (core community extended strategy, CE)

输入 图 $G=(V,E)$, 聚类个数 k , 初始中心社区集合 $C^{(0)} = \{C_1^{(0)}, C_2^{(0)}, \dots, C_k^{(0)}\}$, 其中 $C_i^{(0)} \subseteq V$ 。

输出 社区集合 $C = \{C_1, C_2, \dots, C_k\}$ 。

1) 令集合 $T^{(0)} = V(G) - \cup_{i=1}^k V(C_i^{(0)})$, $\beta_0 = 0.7$, $t=0$ 。

2) 令 $t=t+1$, 对每个社区 $C_1^{(t-1)}, C_2^{(t-1)}, \dots, C_k^{(t-1)}$, 执行下列操作:

① $C_i^{(t)} = C_i^{(t-1)} (1 \leq i \leq k)$, $N_c(C_i^{(t-1)}) = \{u | u \in N_c(x) \wedge x \in C_i^{(t-1)}\}$ 。

② 对任意元素 $v \in T \cap N_c(C_i^{(t-1)}) (1 \leq i \leq k)$, 如果 $b(v, C_i^{(t-1)}) \geq \beta_{t-1}$, 则 $C_i^{(t)} = C_i^{(t-1)} \cup \{v\}$ 。

③ 令 $\beta_t = \beta_{t-1} - 0.1, T^{(t)} = V(G) - \cup_{i=1}^k V(C_i^{(t)})$, 若 $\beta_t \geq 0.3$, 且 $T^{(t)} \neq \emptyset$, 返回步骤 2)。

3) 结束, 输出社区集合 $C = \{C_1^{(t)}, C_2^{(t)}, \dots, C_k^{(t)}\}$ 。

图 3 给出了 BDSG 算法在空手道俱乐部数据集上的聚类结果, 共得到 2 个社区, 空白结点表示未聚类结点。

4 实验与结果分析

为了分析研究 BDSG 算法在真实网络中社区发现的有效性, 将 BDSG 算法分别应用于空手道俱乐部 (Karate)^[23]、海豚社交网络 (Dolphin)^[24]、大学生足球网络 (Football)^[25]、电子邮件网络 (Email)^[26] 和合作网络 (NetScience)^[27] 等 5 个数据集。实验所用计算机配置为 Inter Core i5 CPU 2.5 GHz, 6 GB 内存, Windows 7 操作系统。程序采用 java 编程语言并在 Eclipse 环境下运行。依经验选择密度阈值 $d=0.9$, 调节参数 $\alpha=0.8$ 。

图 3~5 分别给出了本文 BDSG 算法在空手道俱乐部、海豚社交网络和大学生足球网络的聚类结果。表 2 给出了 BDSG 算法与 CPM、 k -dense 算法分别在聚类个数、未聚类结点数、社区模块性 (Q)^[28] 以及运行时间等方面的比较结果。

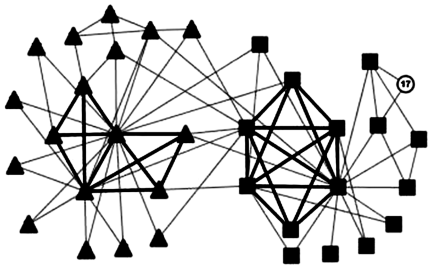


图 3 BDSG 算法在空手道俱乐部得到的聚类结果
Fig.3 Clustering results on zachary's karate club obtained by BDSG

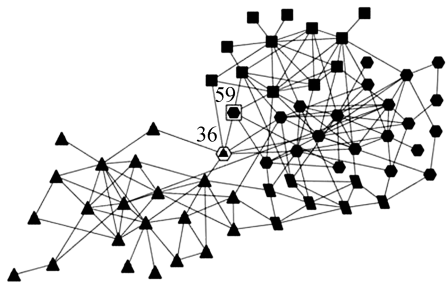


图 4 BDSG 算法在海豚社交网络上得到的聚类结果
Fig.4 Clustering results on dolphins social network obtained by BDSG

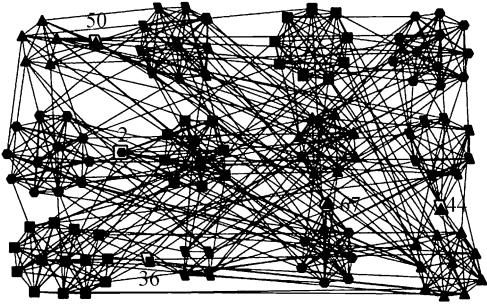


图 5 BDSG 算法在大学生足球网络上得到的聚类结果
Fig.5 Clustering results on college football network obtained by BDSG

表 2 不同数据集上聚类结果的比较

Table 2 The comparison of the clustering results among different datasets

数据集	顶点数	边数	原始社区个数	算法	聚类个数	未聚类节点数	Q	运行时间/ms
空手道俱乐部	34	78	2	BDSG	2	1	0.820 5	93
				CPM	3	22	0.192 3	87
				k -dense	2	22	0.294 8	129
				CPM+CE	3	3	0.410 2	117
				k -dense+CE	2	1	0.820 5	165
海豚社交网络	62	159	2	BDSG	4	0	0.773 5	149
				CPM	4	34	0.408 8	175
				k -dense	4	34	0408 8	568
				CPM+CE	4	16	0.591 1	202
				k -dense+CE	4	16	0.591 1	599
大学生足球网络	115	613	12	BDSG	12	0	0.639 0	480
				CPM	13	2	0.595 1	920
				k -dense	12	2	0.637 0	1 860
				CPM+CE	13	0	0.601 0	1 028
				k -dense+CE	12	0	0.648 0	1 986
电子邮件网络	1 133	5 451	—	BDSG	28	34	0.722 4	60 797
				CPM	55	562	0.268 7	592 410
				k -dense	6	558	0.251 7	55 240
				CPM+CE	55	341	0.289 7	601 835
				k -dense+CE	6	14	0.503 4	63 938
合作网络	1 589	2 742	—	BDSG	134	657	0.782 8	21 273
				CPM	159	843	0.520 1	97 161
				k -dense	91	843	0.730 5	15 352
				CPM+CE	159	688	0.567 5	120 927
				k -dense+CE	91	790	0.763 1	23 451

实验结果表明 BDSG 算法在这些网络数据上均具有较好的性能表现。BDSG 算法在空手道俱乐部和大学生足球网络上所得到社区个数与网络实际的社区个数相同,而电子邮件网络和合作网络缺乏原始社区个数信息,无法进行比较;海豚社交网络 and 大学生足球网络中,BDSG 算法所用时间明显少于 CPM 与 k -dense 算法;在电子邮件网络和合作网络中,BDSG 运行时间比 k -dense 算法慢,但最终未聚类结点数少于 k -dense 算法;在这些实验数据集上,本算法所产生的未聚类结点数明显较少、社区模块性较高。

此外,本文给出的中心社区扩展算法也可应用于 CPM、 k -dense 等算法以处理未聚类节点,提高聚

类性能。实验结果(见表 2)表明 CPM 与 k -dense 算法的聚类有效性均显著提高。在空手道俱乐部、海豚社交网络、电子邮件网络和合作网络中,在 CPM 与 k -dense 算法运行时间略有增大的情况下,CE 算法的加入使得其未聚类结点数降幅较大,社区模块性具有较为明显的提高。同时 CPM 与 k -dense 算法在加入扩展策略 CE 之后与 BDSG 算法相比,BDSG 算法在未聚类结点数以及社区模块性方面优势依然较为明显。

综上所述,BDSG 算法在空手道俱乐部、海豚社交网络、大学生足球网络、电子邮件网络和合作网络等数据集上,与 CPM、 k -dense 算法相比运行时间较短、未聚类结点数较少、社区模块性较高,具有良

好的聚类性能。同时,中心社区扩展算法可以有效地提高 CPM、 k -dense 算法的聚类性能,该算法也可用于其他非结点完全覆盖算法。

5 结束语

本文提出一种基于稠密子图的图聚类算法 BDSG,解决了基于密度算法中大量未聚类结点问题。通过搜索网络中的相对稠密子图得到中心社区;通过定义结点对社区的归属度来度量结点和社区连接倾向性,进而给出一种中心社区扩展策略对中心社区外结点进行聚类。通过与 CPM、 k -dense 算法在 5 个真实网络数据集上进行分析比较,结果表明,BDSG 算法在未聚类结点数、模块性及运行时间方面均表现出较好的性能。同时中心社区扩展策略与其他算法相结合,对提高 CPM、 k -dense 等算法的聚类性能具有一定的适用性。

参考文献:

- [1] FORTUNATO S. Community detection in graphs[J]. Physics reports, 2010, 486(3/4/5): 75-174.
- [2] NEPUSZ T, YU Haiyuan, PACCANARO A. Detecting overlapping protein complexes in protein-protein interaction networks[J]. Nature methods, 2012, 9(5): 471-472.
- [3] DEYLAMI H A, ASADPOUR M. Link prediction in social networks using hierarchical community detection[C]//Proceedings of the 7th conference on information and knowledge technology. Urmia, Iran, 2015: 1-5.
- [4] SCHAEFFER S E. Graph clustering[J]. Computer science review, 2007, 1(1): 27-64.
- [5] 杨博, 刘大有, LIU Jiming, 等. 复杂网络聚类方法[J]. 软件学报, 2009, 20(1): 54-66.
YANG Bo, LIU Dayou, LIU Jiming, et al. Complex network clustering algorithms [J]. Journal of software, 2009, 20(1): 54-66.
- [6] 冀俊忠, 刘志军, 刘红欣, 等. 蛋白质相互作用网络功能模块检测的研究综述[J]. 自动化学报, 2014, 40(4): 577-593.
JI Junzhong, LIU Zhijun, LIU Hongxin, et al. An overview of research on functional module detection for protein-protein interaction networks[J]. Acta automatica sinica, 2014, 40(4): 577-593.
- [7] PALLA G, BARABÁSI A L, VICSEK T. Quantifying social group evolution[J]. Nature, 2007, 446(7136): 664-667.
- [8] SIDIROPOULOS A, PALLIS G, KATSAROS D, et al. Pre-fetching in content distribution networks via web communities identification and outsourcing[J]. World wide web, 2008, 11(1): 39-70.
- [9] 陈克寒, 韩盼盼, 吴健. 基于用户聚类的异构社交网络推荐算法[J]. 计算机学报, 2013, 36(2): 349-359.
- CHEN Kehan, HAN Panpan, WU Jian. User clustering based social network recommendation[J]. Chinese journal of computers, 2013, 36(2): 349-359.
- [10] FREY B J, DUECK D. Clustering by passing messages between data points[J]. Science, 2007, 315(5814): 972-976.
- [11] NEWMAN M E J. Community detection and graph partitioning[J]. Europhysics letters, 2013, 103(2): 28003.
- [12] NEWMAN M E J. Spectral methods for community detection and graph partitioning[J]. Physical review E, 2013, 88(4): 042822.
- [13] LIN Chuncheng, KANG Jiarong, CHEN J Y. An integer programming approach and visual analysis for detecting hierarchical community structures in social networks[J]. Information sciences, 2015, 299: 296-311.
- [14] REN Jun, WANG Jianxin, LI Min, et al. Identifying protein complexes based on density and modularity in protein-protein interaction network [J]. BMC systems biology, 2013, 7(S4): S12.
- [15] LI Xiaoli, FOO C S, NG S K. Discovering protein complexes in dense reliable neighborhoods of protein interaction networks [C]//Proceedings of the computational systems bioinformatics conference. San Diego, USA, 2007, 6: 157-168.
- [16] PALLA G, DERÉNYI I, FARKAS I, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. Nature, 2005, 435: 814-818.
- [17] SAITO K, YAMADA T, KAZAMA K. Extracting communities from complex networks by the k -dense method[J]. IEICE transactions on fundamentals of electronics, communications and computer sciences, 2008, E91-A(11): 3304-3311.
- [18] SUN Penggang, GAO Lin. Fast algorithms for detecting overlapping functional modules in protein-protein interaction networks[C]//Proceedings of the IEEE computational intelligence in bioinformatics and computational biology. Nashville, TN, USA, 2009: 247-254.
- [19] LIU Guimei, WONG L, CHUA H N. Complex discovery from weighted PPI networks[J]. Bioinformatics, 2009, 25(15): 1891-1897.
- [20] BADER G D, HOGUE C W V. An automated method for finding molecular complexes in large protein interaction networks[J]. BMC bioinformatics, 2003, 4(1): 2.
- [21] FREEMAN L C. A set of measures of centrality based on betweenness[J]. Sociometry, 1977, 40(1): 35-41.
- [22] CUI Yaizu, WANG Xingyuan, EUSTACE J. Detecting community structure via the maximal sub-graphs and belonging degrees in complex networks[J]. Physica A: statistical mechanics and its applications, 2014, 416: 198-207.
- [23] ZACHARY W W. An information flow model for conflict

and fission in small groups[J]. Journal of anthropological research, 1977, 33(4): 452-473.

[24] LUSSEAU D, SCHNEIDER K, BOISSEAU O J, et al. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations: Can geographic isolation explain this unique trait[J]. Behavioral ecology and sociobiology, 2003, 54(4): 396-405.

[25] NEWMAN M E J. Finding community structure in networks using the eigenvectors of matrices[J]. Physical review E, 2006, 74(3): 036104.

[26] GUIMERÀ R, DANON L, DÍAZ-GUILERA A, et al. Self-similar community structure in a network of human interactions[J]. Physical review E, 2003, 68(6): 065103.

[27] NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks[J]. Physical review E, 2004, 69(2): 026113.

[28] SHEN Huawei, CHENG Xueqi, CAI Kai, et al. Detect overlapping and hierarchical community structure in networks[J]. Physica A: statistical mechanics and its applications, 2009, 388(8): 1706-1712.

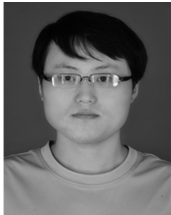
作者简介:



郑文萍, 1979 年生, 女, 副教授, 中国计算机学会会员, 主要研究方向为图论算法、生物信息学等。主持多项国家级项目, 发表学术论文多篇。



张浩杰, 1991 年 8 月生, 男, 硕士研究生, 主要研究方向为数据挖掘、图聚类类。



王杰, 1988 年 8 月生, 男, 博士研究生, 主要研究方向为数据挖掘、生物信息学。

2016 年协同软件过程国际会议

International Track on Collaborative Software Processes

Building complex and trustworthy software systems in the shortest time-to-market remains the challenging objective that competitive companies are facing constantly, and mastering the development process is a key issue towards this objective. Companies and software developers need to be able to formalize their development processes in order to analyze them, to enact them, and to assess them for quality measurement and efficiency improvement. In the context of complex software and systems development, there are many actors involved in the development who use different tools and heterogeneous modeling languages. So we can say that resulting processes are highly collaborative and must be described and enacted as such.

Most published papers propose innovative technical and human approaches to expand collaboration support, often backed up by theory brought from various disciplines including, management science, design science, cognitive sciences and social sciences. The track CSP'2016 mainly seeks papers with theory, models, design principles, methodologies, and case studies that contribute to better understand the complex interrelations between collaboration and technology for software and systems development processes. Considering the heterogeneity of research in collaboration and technology, researchers may address the validation of their work through multiple approaches including laboratory experiments, fieldwork, analytic evaluations, case studies, prototyping, and empirical tests.

The aim of this track is to gather researchers and industrial practitioners working in the field of Collaborative Process Support for Software Development. It will provide an opportunity for the community to exchange ideas and to present emerging new technologies and understanding in the field.

Website: <https://www.irit.fr/CSP2016/>