

染色问题的网络特性^{*})

朱孟潇 蔡庆生

(中国科学技术大学计算机科学技术系 合肥 230027)

摘要 染色问题是约束满足问题的一个经典问题。通过分析染色问题的经典实例,发现染色问题具备复杂网络中常见的“小世界”特性,即染色问题所构成的网络中,任意两个节点之间的平均路径长度很小,整个系统呈现出高聚集度的特性,以及节点度的特异分布。这些特性是随机图所不具备的,因此,以随机图做染色问题测试集的传统方法是不完善的。在实验中同时发现,染色问题中最小染色数的大小与系统聚集度的大小密切相关,随聚集度的增大,呈指数关系增大。

关键词 染色问题,网络特性,小世界,聚集度,最小染色数

Network Characters of Coloring Problems

ZHU Meng-Xiao CAI Qing-Sheng

(Department of Computer Science, University of Science and Technology of China, Hefei 230027)

Abstract Coloring Problem is a classical Constrain Satisfaction Problem. By analyzing the benchmark examples of Coloring Problem, “Small World” characters are found, which are quite common in complex networks. In Small World networks, the average path lengths between each two nodes are small, while the whole system express high cluster tendency and the distributions of node degrees are skew. All these properties are lacked by Random Graphs, and thus we questioned completeness of the traditional method of utilizing random graph as benchmarks. Through the experimental results, the Least Coloring Number is found to have close relationships with the clustering coefficient of the network. With the raise of the clustering coefficient, the least coloring number increases exponentially.

Keywords Coloring problem, Network characters, Small world, Clustering coefficient, Minimal coloring number

1 引言

在许多自然和人造的复杂系统中都存在着网络结构,例如生物网络、社会网络以及约束满足问题所构成的网络等等。它们共同的特性就是,系统由多个结构类似的个体所组成。例如,在约束求解问题中,可以将一个个变量看作是个体并用点来表示,而用边来表示变量之间的联系,那么整个约束求解问题所构成的系统就可以用一个网络^[2]来表示。通过研究其网络特性,发现这些复杂系统的构造通常既不是规则的,也不是完全随机的^[1],而是具备一种“小世界”^[3]的特性,主要包括任意两个节点之间的平均路径长度很短,以及整个系统呈现高度聚集的特性。这种特殊的网络特性将影响系统的性能,以及问题的求解过程。

本文将具体分析经典的约束求解问题——染色问题的图所呈现出来的“小世界”特性,并考察染色问题的网络特性对问题解的影响。本文第2节简单介绍约束满足问题与染色问题定义;第3节介绍小世界特性;第4节介绍通过实验获取的实际染色问题的网络特性,指出利用传统的产生随机图的方式对实际染色问题模拟的不完备性,并分析聚集度特性对最小染色数的影响;最后是结论。

2 染色问题及其网络构成

染色问题(Coloring Problem)是图论和约束求解问题中的一个经典问题。首先介绍一下约束满足问题。约束满足问题(Constraint Satisfaction Problems, CSPs)^[4]可以涵盖计算机科学领域当中的一大类问题,例如调度问题、计划问题、设

计问题、资源分配、图形问题、诊断问题、配置问题、定性推理、实时系统、机器人、视觉问题、用户模型、分子生物等。自20世纪70年代起,来自人工智能、程序语言、符号计算和可计算逻辑等领域的研究者都对约束满足问题领域进行了大量的研究。约束满足问题由一个有限的变量集合,一个所有变量的有限离散值域集合及一个约束关系的集合组成。

定义1 约束满足问题 P ,包括

1. 一个变量的有限集合, $X = \{X_1, X_2, \dots, X_n\}$;
2. 一个值域的集合 D ,包含所有变量的有限离散值域, $D = \{D_1, D_2, \dots, D_n\}$, 对 $\forall i \in [1, n], X_i \in D_i$;
3. 一个约束关系的集合, $C = \{C(R_1), C(R_2), \dots, C(R_m)\}$ 。 R_i 是第 i 个约束条件,是变量集的一个有序子集, $C(R_i)$ 是 R_i 中所有变量的笛卡尔乘积的子集。不满足约束条件又称为“约束冲突”。

定义2 一个解答 S 是一组所有变量的赋值,并且满足所有的约束条件。

1. S 是一个有序集, $S = \langle v_1, v_2, \dots, v_n \rangle, S \in D_1 \times D_2 \times \dots \times D_n$;
2. $(\forall j \in [1, m])(\exists S' \subseteq S) \wedge (S' \models C(R_j))$ 成立,且 S' 也是一个有序集。

染色问题在现实生活中也有着很多的应用,例如在学生课表编排、电网分布、计算机路由分布等等实际问题当中。它要求用 m 种颜色对一个图的所有顶点染色,使每条边相连的2个顶点颜色互不相同。对于染色问题,等价的约束满足问题为:一个顶点的颜色由一个变量表示,每个变量的值域为 m 种颜色,约束关系是一个2元关系,即每2个有边相连的顶点

^{*})本文受国家自然科学基金资助(No. 70171052)。朱孟潇 硕士生,主要研究领域为人工智能,多主体系统,复杂性系统;蔡庆生 教授,博士生导师,主要研究领域为人工智能,机器学习,多主体系统。

的相关变量不能取相同的值。图 1 是一个简单例子: 变量集合 $X = \{V_1, V_2, V_3, V_4\}$, 变量值域分别为, $D_1 = \{\text{green}, \text{red}\}$, $D_2 = \{\text{red}, \text{blue}\}$, $D_3 = \{\text{blue}\}$, $D_4 = \{\text{blue}, \text{green}\}$, 约束是 $V_1 \neq V_2, V_1 \neq V_3, V_1 \neq V_4$ 和 $V_2 \neq V_3$, 所以约束集合 $C = \{\langle \text{green}, \text{red} \rangle, \langle \text{green}, \text{blue} \rangle, \langle \text{red}, \text{blue} \rangle, \langle \text{green}, \text{blue} \rangle, \langle \text{red}, \text{blue} \rangle, \langle \text{green}, \text{blue} \rangle, \langle \text{red}, \text{blue} \rangle, \langle \text{red}, \text{green} \rangle, \langle \text{red}, \text{blue} \rangle\}$ 。

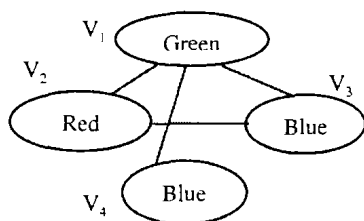


图 1 染色问题的一个例子

约束满足问题可以很容易地抽象成为一个网络, 将每个变量当作网络中的一个节点, 变量之间的约束关系构成各个边, 从而构成一个网络。染色问题对应的网络则为待染色的图。

3 小世界特性

“小世界”特性是指网络的拓扑结构上呈现出局部高度聚集的特性, 而从整体上任意两个节点之间的路径相对来说很短的^[5]。这种特殊的性质在实际的复杂系统当中非常常见。例如在神经网络^[6], 飞机航班的时刻表^[7], 血管的分布^[8-9], 公司之间合作的关系^[10], 人们参加相同社会活动的网络^[11]等等网络当中, 都可以看到小世界性质的存在。

具有小世界特性的网络模型所具有的三个主要的特性: 短路径特性、高聚集度特性和节点度的分布特性。

3.1 短路径特性

在一个连通图中, 任意两个节点之间都存在一条或多条路径。定义任意两个节点之间最短路径的平均值 L 为整个网络的特征路径长度。对于包含多个节点的大规模网络, 一般直观地会认为 L 值将很大, 但是事实恰恰相反, 在一个有着成千上万个节点的图中, 特征路径 L 的值反倒会非常小。如表 1 所示, 可以看到四个网络的例子^[12]。其中, 第一个例子中, 节点为来自 1999 年的 Fortune 1000 中公司的领导者, 若两个领导者在同一家公司的董事会中工作, 那么代表这两领导者的两个节点间就存在一条连线相连。类似地, 第二个例子中的数据来自 Internet Movie Database (截至 2000 年 5 月 1 日的数据), 如果两个演员在同一部电影当中出现, 那么这两个演员

之间就认为是连接的。第三个例子中的数据来自 Los Alamos E-print Archive 当中的论文记录, 如果两个科学家在同一篇论文当中出现, 则代表着两个科学家的节点之间就有一条路径相连, 类似地, 构成了第四个例子中的生物论文作者组成的网络。从表中的数据可以看到, 对于包含有 N 个节点 ($N \gg 1000$) 的真实系统, 平均的最短路径长度都是非常小的, 它们都处于 3~6 之间。也就是说任何两个节点之间都可以通过较短的路径相互到达。

表 1 短路径特性及高聚集度的体现

网络	来源	大小	L	C
公司领导者	Fortune	7,673	4.60	0.588
电影演员	IMDB	449,913	3.84	0.199
物理论文作者	Arxiv.org	52,909	5.92	0.452
生物论文作者	Medline	1,520,253	4.42	0.088

3.2 高度的聚集特性

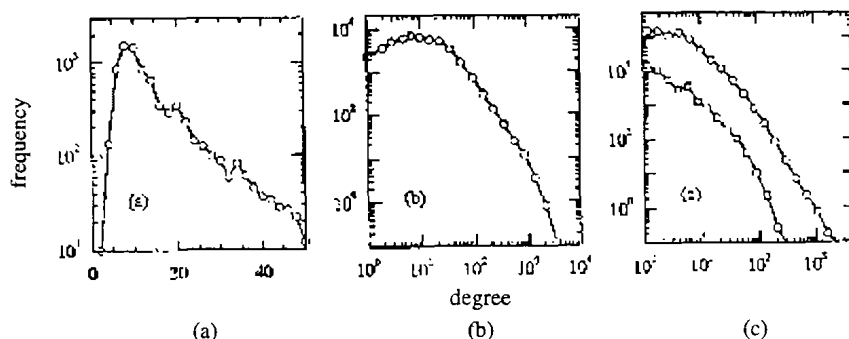
“小世界”现象所具备的另外一个性质就是高度的聚集特性^[2,13]。以物理学论文作者网络为例, 就是说在两个分别与同一个研究者合作的人相互之间合作的比率是非常大的, 远高于随机选取的两个研究者。定量地, 可以这样计算聚集度 (Clustering Coefficient) C 的值^[12,13]:

$$C = \frac{3 \times \text{图中三角形的数目}}{\text{三关联节点的数目}} \quad (1)$$

这里“三角形”指的是三个互相之间有边相连的节点对的数目, 而“三关联节点”是指任意一个三节点对满足至少一个与另外两个相连的条件。公式当中乘 3 的系数是由于每个三角形都应当在三个节点处分别计算一次, 这就保证聚集度 C 的值就是一个 0 到 1 之间的数。那么, 根据定义, 在一个全连通图当中, 由于任意两个节点之间都有边相连, 这时 $C=1$ 。在上面所举的例子当中, 如表 1 所示, 可以看到, 聚集度的值是一个小于 1, 但比随机的可能性 $O(N^{-1})$ 大很多的一个值。

3.3 节点度的分布特性

对于一个网络所组成的图, 根据图论当中的定义, 将与一个节点相连的边的数目定义为该节点的度。这样, 就可以对各个节点的度进行统计, 进而对节点度的分布进行统计, 并从统计学的角度考察分布的特性。由于被考察的系统并不是从一开始构建的时候就有某种规则指引, 所以一个直观的猜测是这些度的分布可能是日常生活中常见的正态分布或是泊松分布。但是事实却不是这样。这些自组织的系统具备一些特殊的性质, 有些图的度是按照指数分布的, 而更多的则是按照幂次率分布的。同样对于上面的四个例子, 可以看到它们的度分布的规律图^[12,14], 如图 2 所示。



(a) 为公司领导者的度的分布; (b) 为电影演员度的分布;

(c) 上面一条曲线为生物论文作者度的分布, 下面的曲线为物理论文作者度的分布

图 2 节点度的分布图

从图 2 中可以看到, 除了第一个公司领导者的例子中, 节点的度基本符合指数分布之外, 另外三个系统当中, 度的分布

都是基本符合幂次率的分布的。也就是说,这些网络表现出一些标度无关的特性。这些节点度的分布,无疑也会对整个系统的性质产生很大的影响。

4 染色问题的网络特性

本文分析了 DIMACS (Center for Discrete Mathematics and Theoretical Computer Science, <http://dimacs.rutgers.edu/Challenges/Seventh/#PC>) 上 Donald Knuth 的斯坦福大学经典大规模染色问题。从中发现,其中几乎所有的图都具备前面所描述的短路径、高聚集度以及节点度的分布等等特征。本文考察了以下的染色实例:

1. anna.col (138,496), 11, SGB
2. david.col (87,406), 11, SGB
3. miles1000.col (128,3216), 42, SGB
4. fpsol2.i.1.col (496,11654), 65, REG
5. fpsol2.i.2.col (451,8691), 30, REG
6. mulsol.i.1.col (197,3925), 49, REG
7. mulsol.i.2.col (188,3885), 31, REG
8. zeroin.i.1.col (211,4100), 49, REG

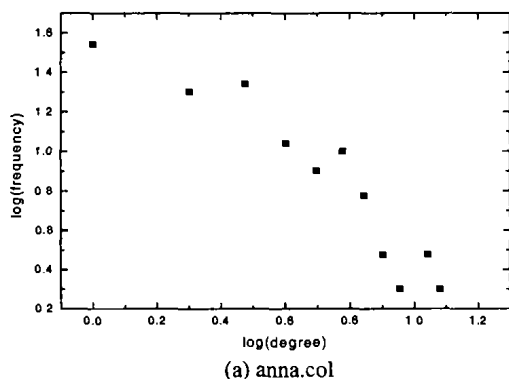
其中,* .col 为文件名,括号中的数字分别为节点的数目和路径的数目。括号后的数字为染色需要的最少颜色,最后的三个字母为文件的来源。

4.1 短路径与高聚集度特性

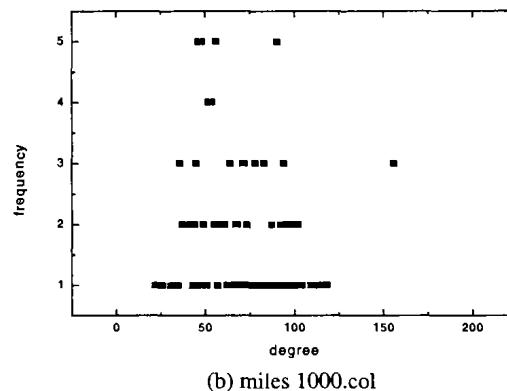
通过计算,可以得出根据实际网络计算出来的平均最短路径 L 和聚集度 C 的值,与随机图的相应数据的对比关系。如表 2 所示。可以看出,一方面来讲实际网络中的平均最短路径

表 2 DIMACS 中实际染色问题与随机图中平均最短路径和聚集度的对比

染色问题	L	L_{random}	C	C_{random}
anna.col	2.438	2.506	0.268	0.0518
david.col	1.935	1.999	0.351	0.107
miles1000.col	1.917	1.239	0.756	0.393
fpsol2.i.1.col	1.672	1.612	0.538	0.0947
fpsol2.i.2.col	1.865	1.674	0.226	0.0855
mulsol.i.1.col	1.580	1.434	0.747	0.202
mulsol.i.2.col	1.735	1.407	0.498	0.220
zeroin.i.1.col	1.473	1.462	0.699	0.184



(a) 为 anna.col 中节点度的分布



(b) 为 miles1000.col 中节点度的分布

图 4 节点度的分布图

如图所示,可见实际染色问题中节点度的分布也不是完全随机的,而是符合一定的规律的。图 4(a)为根据实验数据得出的 anna.col 中节点度的分布。其中,横坐标为节点度取

径 L 与随机图 L_{random} 的差别并不是非常大,相对于网络中的节点数目都非常小。在这一点上,网络都体现出了短路径的特性。另外,对于聚集度来说,随机图与实例网络有着很大的差别。比较四、五两列数据可以发现,随机图的聚集度值 C_{random} 远远小于实际图中的 C 值。这一点,就从另外一个方面体现了染色问题的实例网络符合高聚集度的性质。因此具有一种特殊的结构特征。这也就证明了利用随机图作为染色问题测试集的不完备性,因为实际的染色问题具有完全不同的网络特性。而这一特性将对染色问题的求解产生很大的影响。

4.2 聚集度对最小染色数的影响

染色问题的最小染色数定义为满足问题约束所需要的最少颜色数目。而聚集度的大小对最小染色数有着明显的影响。对于全连通图,这时聚集度 $C=1$,而最小染色数为节点数目 N ;而对于完全离散的图,网络退化为单个的节点,最小染色数为 1。为考察染色问题中聚集度与染色数的关系,并排除由于节点数目增加对于染色数的影响,这里将染色数除以节点数目进行归一化。则 $C=1$ 时,归一化后的染色数为 1;而 $C=0$ 时,当 $N \rightarrow \infty$,归一化后的染色数趋近 0。这就将染色数影射到 $(0,1]$ 的区间上。计算结果如图 3 所示,随聚集度的增大,染色数呈现指数增长的趋势。即随聚集度增加,需要的最小染色数急剧增加。

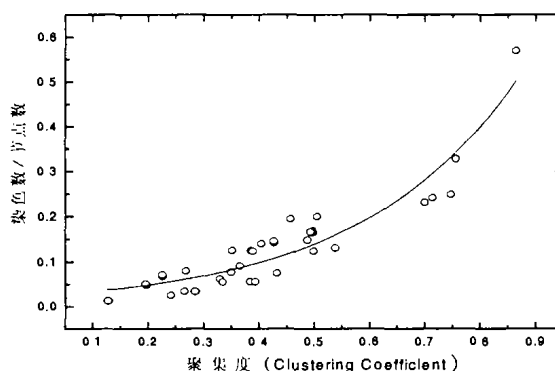


图 3 聚集度对最小染色数的影响

4.3 节点度分布规律

染色问题的大部分实例所构成的网络的节点度分布都存在一些规律。其中较具代表性的例如图 anna.col 和 miles1000.col 中节点度的分布情况,如图 4 所示。

log 值之后的数据,而纵坐标为对应的度出现频率取 log 值之后的数据。在双对数坐标下,可以看到,度的分布几乎呈一条

(下转第 32 页)

变。这种技术中,本地机收到远端代码前,要求节点管理者注册公钥到本地数据库。需要注意的是即使移动代码是可信的,并不意味着远端代码是无害的。

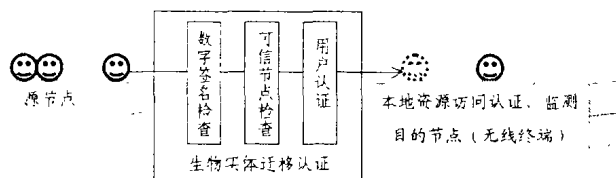


图4 无线安全认证

JVM 为程序执行创建了保护域,这限制了远端代码权利。一个生物实体迁向新的节点时,节点上的 Java 类加载器检查移动代码的创建者。如果创建者不在本地列表中,JVM 用最少的权利执行代码。JVM 检查到代码的来源前,应在本地机上注册可信的节点。

生物网络平台系统初始化时,要求注册一个可信的服务列表,功能是管理生物实体对于本地资源的访问控制。实体属性和用户认证与安全信息比较来形成访问控制列表。生物网络容器安全认证服务完成这些认证,包括完成生物实体向本地资源的访问控制。生物实体属性和用户认证与本地安全信息比较来形成请求迁移的生物实体的访问控制列表(列表不针对已经在本地节点上的生物实体)。当终端生物网络平台收到未知节点上实体的迁移请求,或收到一个无标记的生物实体时,安全认证服务会向用户发出信号,然后由终端用户决定是否接受这个迁移请求。如果得到用户的允许,生物实体向终端迁移。基于用户的决定,安全认证服务向不同的生物实体提供不同的访问权利。

生物网络服务器中的本地安全服务则创建了内部列表,它包含资源名称和允许的操作。资源列表与每个保护域相关,一个生物实体访问本地资源时,本地安全服务将检查内部列

表中是否有请求的资源。它还提供了监测资源访问的功能,一旦生物实体迁移到用户的终端设备中,本地安全服务就监测资源访问,保护事先定义的访问控制列表。如有生物实体非法访问资源时,用户或者拒绝这个访问,或者修改访问控制列表,让这些资源能被访问。本地安全服务包含本地资源访问认证的密码,密码存储在终端设备中,认证也在设备中完成。它扩展了 Java 默认安全机制,利用 Java 内在机制实现密码认证。如果终端用户想禁止从某个节点迁移来的生物实体,它可以拒绝所注册的该节点的请求。

结束语 生物网络架构和移动 Agent 框架^[3]中的实体都具有移动性。移动 Agent 框架允许 Agent 从一个网络节点迁移到另一个网络节点,搜索并与服务交互。移动 Agent 是自治的,它们可以根据当前状态进行自主迁移,但这种自治性仅限于迁移行为,仅仅通过对象迁移所提供的适应性是非常有限的。而能自治地完成许多行为的生物实体可以获得很强的自适应能力去迎接未来 Internet 高度异构和动态的网络计算环境所带来的挑战。另外,移动 Agent 框架中一些实现定义了对应的最大消耗资源量,而生物网络架构以能量为主线的控制策略允许生物实体对其自身的行为负责,可以突现出更多种类、更能满足用户需要的服务。

参考文献

(上接第 25 页)

直线。也就是说,节点的度符合幂次率的分布。直观地说,也就是图中,度较小的节点出现的频率相对较大,而度较大的节点出现的频率相对较小。(b)中,为另外一个图 miles1000.col 中节点度的分布,其中,横坐标为节点的度,纵坐标为节点度出现的频率。从分布图可以看出,节点度的分布是相对均匀的。度较高的节点出现频率也有很高的,而度较低的节点出现频率也有很低的。也就是说,节点分布较为均匀。

结论 对于包含多个约束变量的约束满足问题,可以利用网络来表示变量之间相互约束的关系。同时系统求解的可行性以及求解的效率无疑会受到网络性质的重大影响。通过对实际染色问题所对应网络的研究,发现大部分的系统都体现出“小世界”的特性,即短路径、高聚集度和节点度的规则分布等性质。并且,计算表明染色问题所需要的归一化的最小染色数与聚集度的大小呈指数关系。这些性质的发现表明,传统的通常采用随机图的方式产生测试集的方法是不完善的。并且,许多传统的搜索算法都是基于随机图的,因此这些搜索求解算法对于这些实际的具有“小世界”特性的例子求解的效果可能并不十分令人满意。

参考文献

- 1 Strogatz S H. Exploring complex networks. *Nature*, 2001, 410: 268~276
- 2 Walsh T. Search in a small world. In: *Proc. of IJCAI-99*, 1999
- 3 Watts D J, Strogatz S H. Collective dynamics of 'small-world'

- networks. *Nature*, 1998, 393: 440~442
- 4 Kumar V. Algorithm for Constraint Satisfaction Problem: A Survey. *AI Magazine*, 1992, 13(1): 32~44
- 5 Broder A, Kumar R, Maghoul F, et al. Graph structure in the web. *Computer Networks*, 2000, 33: 309~320
- 6 White J G, Southgate E, Thompson J N, Brenner S. The structure of the nervous system of the nematode *C. elegans*. *Phil. Trans. R. Soc.*, 1986, 314: 1~340
- 7 Amaral L A N, Scala A, Barthelemy M, Stanley H E. Classes of small-world networks. In: *Proc. Natl. Acad. Sci.* 97, 2000. 11149~11152
- 8 West G B, Brown J H, Enquist B J. A general model for the origin of allometric scaling laws in biology. *Science*, 1997, 276: 122~126
- 9 Banavar J R, Maritan A, Rinaldo A. Size and form in efficient transportation networks. *Nature*, 1999, 399: 130~132
- 10 Powell W W, Koput K W, Smith-Doerr L. Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, 1996, 41: 116~145
- 11 Davis A, Gardner B B, Gardner M R. Deep South. University of Chicago Press, Chicago, 1941
- 12 Newman M E J, Strogatz S H, Watts K J. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* 64, 026118, 2001
- 13 Newman M E J. Models of the Small world. Available as cond-mat/0001118
- 14 Newman M E J. The structure and function of networks. on press by Computer Physics Communications
- 15 Dorigo M, et al. The ant system: Optimization by a colony of co-operating agents. *IEEE Transactions on Systems, Man and Cybernetics-Part B*, 1996, 26(1): 1~13
- 16 韩婧,蔡庆生. AER 模型中的涌现智能. 模式识别与人工智能, 2002, 15 (2): 134~142