

基于数据的任务规划在复杂网络中的应用^①

董 玮, 毕于慧, 闫丽丽, 张 权

(61139 部队, 北京 100091)

摘 要: 复杂网络中目标任务的变化能够引起网络的自身演化, 如何对任务负载进行平衡是提高网络自身性能的关键问题之一, 以区域快递网络为研究对象, 提出了基于动态数据和静态数据对对象进行配置、拆分、添加和抛弃的负载平衡规则, 研究了其在任务规划中对网络负载平衡起到的作用, 实验结果表明, 提出的规则能够有效提高类似网络的稳定性和网络资源的合理配置.

关键词: 复杂网络; 任务规划; 静态数据; 动态数据

Application Study of Data-Based Task Planning in Complex Networks

DONG Wei, BI Yu-Hui, YAN Li-Li, ZHANG Quan

(61139 Troop, Beijing 100091, China)

Abstract: The task changes of Complex Network can cause the evolvement of the network, one of the key problems is how to balance the task load to improve the network's performance. The load balancing rules to configure, divide, add and abandon the object were proposed based on the static and dynamic data with regional express network used as the research object. And then the action of the rules to the load balancing in the task planning was studied. The experimental results show that the rules proposed can effectively improve the stability of similar network and promote the rational use of the network resources.

Key words: complex networks; mission planning; static data; dynamic data

① 收稿时间:2014-09-22;收到修改稿时间:2014-11-06

1 引言

复杂网络是近年来国内外学者研究的一个热点问题, 在现实社会中, 许多真实的系统都可以用网络来表示. 例如, 万维网可以看作是网页之间通过超级链接构成的网络, 社会关系网络可以看成是由不同人群的合作关系构成的网络, 基因调控网络可以看作是不同的基因通过调控与被调控关系构成的网络. 复杂网络理论正是通过对网络拓扑结构特征量统计规律的研究来揭示各种复杂网络中共性的网络随机演化性质、结构稳定性及受到攻击时的动力学传播特性. 随着大数据时代的到来, 数据开始反映并成为促成网络变化的因素之一, 研究数据和网络的相互关系, 以及网络节点数据对网络变化的影响对于揭示数据时代的复杂网络的特性及其演变规律具有重要意义^[1,2]. 本文以区域快递网络为对象, 研究了当目标任务的变化引起复杂网络自身演化时, 如何利用数据与网络之间的相互

影响对任务规划负载进行平衡, 以提高类似复杂网络的自然收敛速度, 并提升对应的网络资源配置效率.

2 相关概念

2.1 聚集系数

聚集系数用来描述网络中节点的聚集情况, 即网络有多紧密, 比如在社会网络中, 你朋友的朋友可能也是你的朋友或者你的两个朋友可能彼此也是朋友. 其计算方法为: 假设节点 i 通过 k_i 条边与其它节点相连接, 如果这 k_i 节点都相互连接, 它们之间应该存在 $k_i(k_i-1)/2$ 条边, 而这个节点之间实际存在的边数只有 E_i 的话, 则它们之间的比就是节点聚集系数, 网络的聚集系数就是整个网络中所有节点的聚集系数的平均值^[3].

2.2 动态数据

网络中的动态数据是指在网络中由于节点变动、

节点需求变动以及节点之间关联程度的变化所产生的数据。

与表示系统的成功实验也证明了本文提出的 Web 信息抽取算法的实用性。

2.3 静态数据

网络中的静态数据是指用于描述网络中节点和节点之间关系的固有的、相对稳定的属性数据。

3 区域快递对象属性描述

下面以区域快递作为任务对象, 研究静态数据与动态数据的联合应用对复杂网络任务规划中负载加速平衡的影响。

3.1 任务对象描述

假设区域快递任务中包括: 货物、区域仓库、区域用户、区域快递员等实体对象。我们仅以它们的部分属性作为研究重点, 具体描述如下:

1) 货物(种类、数量、体积、重量)

$$S(\text{sort}, \text{number}, \text{volum}, \text{quality})$$

2) 区域仓库(货物)

$$D(s_1, s_2, s_3, \dots, s_n)$$

3) 区域用户(用户)

$$U(b_1, b_2, b_3, \dots, b_n)$$

4) 区域快递员(快递员)

$$Ex(r_1, r_2, r_3, \dots, r_n)$$

5) 快递时间

$$T(t)$$

6) 区域网络

$$NT(D_i, U_i, Ex_i, T)$$

3.2 区域快递中静态数据与动态数据描述

下面我们对区域快递中存在的静态数据与动态数据的部分属性进行描述, 其中静态数据以 S 标识, 动态数据以 D 标识。

1) 区域仓库数据集(仓库位置量, 容积, 货品,

$$StorageD(\text{Location}(S), \text{Cubage}(S), S(D), \dots)$$

2) 区域用户数据集(用户住址, 购物时间, 货物, 购物数量, ...)

$$UserD(\text{Address}(S), \text{Time}(S), S(D), \text{Number}(D), \dots)$$

3) 区域快递员数据集(人员数量, 熟练度, 快递区

域, 承载重量, 承载体积)

$$ExpressD \left(\begin{matrix} \text{Number}(D), \text{Skill}(D), \text{Area}(S), \\ \text{Load}(S), \text{Room}(S) \end{matrix} \right)$$

3.3 区域快递中任务的描述

1) 用户购物

$\text{Sale}(U_i, S, T)$, 代表用户 U_i 购买货物 S 并要求在 T 时送达。

2) 货物出入库

$\text{Express}(D, S_i, Ex_i)$, 代表从区域仓库 D 中由快递员 Ex_i 负责投送货物 S_i 。

3) 货物快递

$\text{Express}(Ex_i, S_i, U_i)$, 代表快递员 Ex_i 将货物 S_i 投送至用户 U_i 。

4 基于数据的任务规划应用

4.1 任务聚类划分

4.1.1 富节点

网络中少量的节点具有大量的边, 这些节点也称为“富节点”^[4]。它们倾向于彼此之间相互连接, 构成“富人俱乐部”。可以用富人俱乐部连通性 $\Phi(r/N)$ 来刻画这种现象, 它表示的网络中前 r 个度最大的节点之间, 实际存在的边数 L 与这 r 个节点之间总的可能存在的边数 $r(r-1)/2$ 的比值, 即:

$$\phi(r/N) = \frac{L}{r(r-1)/2} = \frac{2L}{r(r-1)}$$

AS 层面网络的富人俱乐部连通性如下图所示:

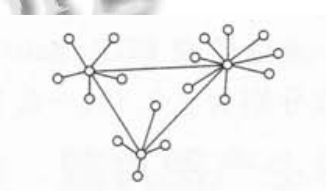


图 1 富节点彼此之间相互连接

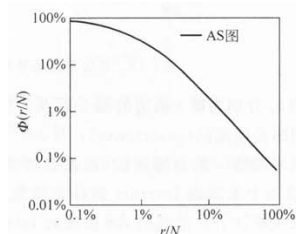


图 2 AS 层面网络的富人俱乐部现象

在区域网络中, 显然区域快递员是“富节点”, 它

负责连接在某个聚类区域内的购物用户, 并和其他快递员发生关联。

4.1.2 群聚属性

从网络的传递属性当中可以清楚的看出与随机行为的偏离情况, 该属性有时也称为群聚属性。在很多网络中, 存在下述情况, 即如果顶点 A 与 B 相连, 并且顶点 B 与 C 相连, 那么顶点 A 也极有可能与 C 相连。从网络中提取群聚结构的方法是聚类分析, 有时也称为等级群聚。在以区域快递任务为例中, 我们以基于地址群居的用户聚集看作群聚, 网络中区域快递员作为“富节点”连接群聚用户^[5,6]。即指, 每个快递员负责一个用户群聚, 这个群聚是由地理位置相互较近的用户组成的。

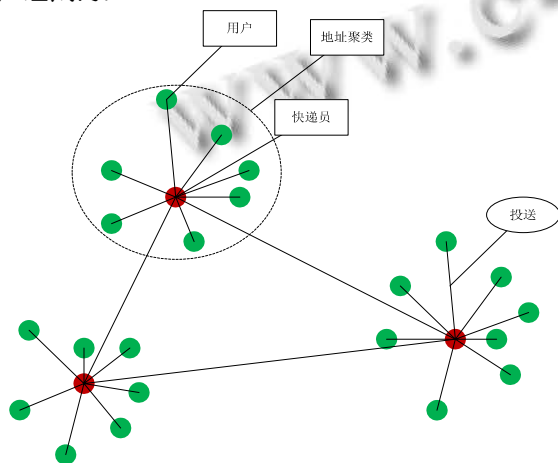


图 3 区域快递网络中的购物用户群聚

4.2 区域快递任务负载均衡规则

区域快递任务负载均衡是指区域仓库满足区域用户当日购买货物种类和数量, 并能够由现有数量的快递员在时限内送达。

4.2.1 货物加权系数

综合考虑货物的在任务中的衍生属性, 如送达时限等, 值越高表示送达时限越短。

则设: $1 \leq \lambda \leq 5$, 即将加权系数分为 5 级。

4.2.2 区域快递员负载

保证快递员当日需投送货物满足自身投送负载, 即

$$\begin{aligned} \text{Storage}(D, S_i, E_{x_i}) &\rightarrow \text{Express}(E_{x_i}, S_i, U_i) \\ \text{Express}(E_{x_i}, S_i, U_i) &\leftrightarrow \text{Sale}(U_i, S_i, T) \\ E_x(L) &\geq \sum_{i=1}^n S_i(q_i)\lambda_i \end{aligned} \quad (1)$$

$$E_x(R) \geq \sum_{i=1}^n S_i(V_i)\lambda_i \quad 1 \leq \lambda_i \leq 5 \quad (2)$$

4.2.3 均衡规则

(1) 节点配置规则

以货物送货单为依据, 每个快递员作为富节点, 连接距离相近用户所产生的聚类中的节点。其中聚类的个数不能超过快递员的数量。在用户加入聚类的过程中:

① 若当前聚类个数小于快递员的数量, 且用户不满足已有聚类规则时, 可建立独立的聚类。

② 若当前聚类个数等于快递员的数量, 则该用户应按距离远近分配至最近的 1 个聚类当中。

(2) 节点拆分规则

当节点配置好后, 需要利用富节点负载平衡公式(1)(2), 对每个聚类中富节点和用户节点的分配情况进行检查。若聚类中用户节点的需求超过富节点的负载, 则需要对现有聚类进行拆分。

聚类拆分规则: 若通过聚类拆分能够使网络负载达到均衡, 则进行简单的聚类拆分。其方法为:

① 从负载超标的聚类中, 找出拆分后能够满足该聚类负载的用户节点。

② 按聚类相近的原则, 将其迁移到最近的聚类中, 即同当前富节点距离最短的富节点上, 若该富节点所在聚类满足负载平衡, 则网络负载平衡。

③ 若距离最近的富节点上不满足负载平衡, 则以此类推, 向下一个距离较近的富节点上迁移, 直到网络达到负载平衡。

$$\begin{aligned} \text{Express}(E_{x_i}, S_i, U_i) \cup \text{Sale} &\rightarrow \\ \text{Express}_j(E_{x_{ij}}, S_i, U_i) \cup \text{Express}_k(E_{x_{ik}}, S_j, U_i) \\ \text{Express}_j(E_{x_{ij}}, S_i, U_i) \cup \text{Sale} &\rightarrow \text{Cluster}_j \\ \text{Express}_k(E_{x_{ik}}, S_i, U_i) \cup \text{Sale} &\rightarrow \text{Cluster}_k \end{aligned}$$

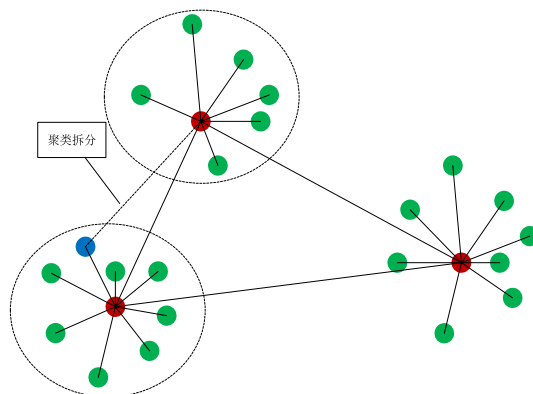


图 4 聚类拆分规则

用户节点拆分规则: 当聚类拆分在整个快递网络中不能满足网络负载时, 需要对用户节点拆分. 其方法为:

① 选择聚类中负载较大的单个用户节点, 对其进行拆分. 该用户节点应靠向另 1 个聚类, 该聚类当前未满载.

② 拆分后, 两个聚类应均达负载平衡.

③ 若较近的聚类均已满载, 则向更远的聚类上迁移拆分后的用户节点, 直到负载平衡.

$$Express(E_{x_i}, S_i, U_i) \cup Sale \rightarrow$$

$$Express_j(E_{x_i}, S_i, U_{ij}) \cup Express_k(E_{x_i}, S_j, U_{ik})$$

$$Express_j(E_{x_i}, S_i, U_{ij}) \cup Sale \rightarrow Cluster_j$$

$$Express_k(E_{x_i}, S_i, U_{ik}) \cup Sale \rightarrow Cluster_k$$

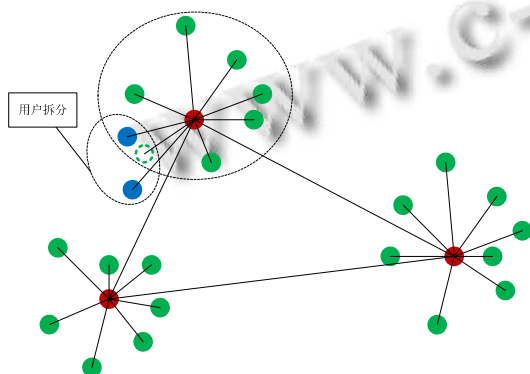


图5 用户节点拆分规则

(3) 富节点添加规则

当以上两种方法均不能满足网络负载平衡时, 则需要使用富节点添加规则. 其方法为:

① 添加富节点, 该富节点的能力应能满足网络负载条件.

② 添加的富节点对已有用户节点进行负载均衡, 其选择的聚类应为负载超标严重的, 保证均衡后, 新聚类能够承担部分其它少量超负载聚类的工作.

$$Ex_{new} \Rightarrow Cluster_{new}$$

$$Express(E_{x_{new}}, S_{ij}, U_{ij}) \cup Sale \rightarrow Cluster_{new}$$

$$Express(E_{x_i}, S_{ik}, U_{ik}) \cup Sale \rightarrow Cluster_{old}$$

$$U(b_i) \cup Sale \rightarrow Cluster_{old}$$

(4) 节点抛弃规则

当拆分规则不能满足网络负载, 同时富节点添加受限的情况下, 应按用户节点的优先级(即加权系数)对已有用户节点进行抛弃操作. 其方法为:

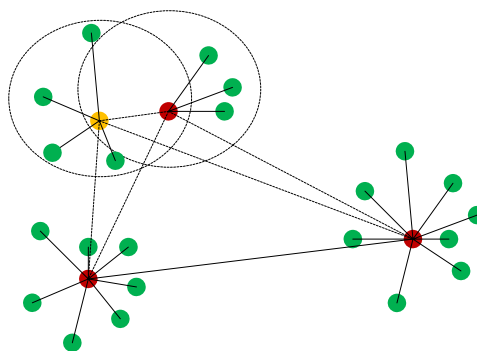


图6 富节点添加规则

① 选取网络中优先级最低的用户节点进行抛弃.

② 当抛弃优先级最低用户节点不能满足网络负载平衡时, 则应按优先级排序情况向上追溯, 直到找到合适的用户节点为止. 例如:

$$U_i \text{ in } Sale(U_i, S_i, (T \geq \kappa))$$

$$\notin \bigcup_{i=1}^n U_i \text{ in } (Express(E_{x_i}, S_i, U_i) \cap Cluster_i)$$

κ 代表货物送达时限门限值, 在抛弃节点时, 首选对货物送达时间较为宽松的节点.

$$\kappa = \frac{(\sum_{i=1}^n \sum (T_i - 24) \text{ in } Sale(U_i, S_i, (T \geq \kappa)))}{n}$$

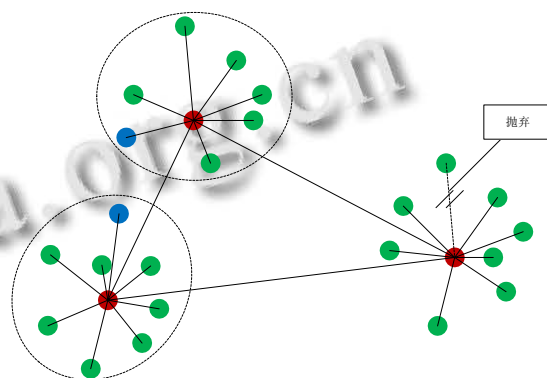


图7 节点抛弃规则

(5) 拓展规则

① 增减富节点数量. 根据区域快递网络中动态数据的积累分析, 可以得出增加快递员即富节点的趋势, 结合每年用户购物习惯、季节变化和购物促销时间, 有针对性的配置区域内快递员数量.

② 提升富节点质量. 有针对性的提升每个快递员的业务水平, 即提高富节点承担负载的能力, 例如: 加强快递员对道路的熟悉程度; 为快递员配置容积更

大、速度更快的送货车辆等。

③ 优化仓库贮物配置方案。对仓库中货物的种类、数量,根据其保质期、货源量等进行合理配置,防止货物在仓库中的积压,保证货物的快速流通。

4.3 结论

通过建立上述基于数据的复杂网络任务负载平衡规则,为特定网络任务需求提供了有效的任务规划。在设置某个快递区域中有 200 用户,6 个快递员负责快递服务(其中 1 个快递员为备份),每个快递员每天工作 12 小时,合计里程为 240 公里,可负载 300 千克货物时,若每个用户(用户间距不大于 10 公里)1 个月内平均 3~5 天产生 1 次 5~20 千克货物快递需求的情况下,图 8 描述了采用本文提出的任务规划方法后,区域快递网络的稳定性和效率都有所提升,复杂网络的收敛速度提高了,对应的网络资源配置效率提升时间得到缩短。

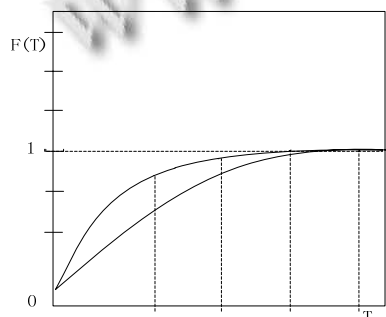


图 8 网络稳定性收敛和效率提升

5 结语

大数据时代数据的运用对研究复杂网络内部的特性和演变规律具有重要意义,本文以区域快递网络为研究对象,以网络内部的静态和动态数据为基础,建立了任务规划负载平衡规则,相类似的复杂网络有很多,例如:银行网点、桶装水送水站、超市、连锁店、抢险救灾等,甚至目前被关注的多源数据中心建设也同其相象,对于提升类似复杂网络的特性,本文的方法有一定的参考价值。

参考文献

- 1 方锦清,汪小帆,刘曾荣.略论复杂性和非线性复杂网络系统的研究.科技导报,2004,(2):9-12.
- 2 许国志等.系统科学.上海:科技教育出版社,2000.
- 3 Han JW, Kamber M. Data Mining Concepts and Techniques, 北京:机械工业出版社,2007.
- 4 Wang XF. Complex networks: Topology, dynamics and synchronization. Int. J. Bifurcation and Chaos, 2002, 12(5): 885-915.
- 5 Li X, Chen GR. A local-world evolving network model. Physica A, 2003, 328: 274-286.
- 6 Albert R, Barabási AL. Topology of evolving networks: Local events and universality. Phys. Rev. Lett, 2000, 85(24): 5234-5237.
- 7 Yook SH, Jeong H, Barabási AL. Weighted evolving networks. Phys Rev. Lett, 2001, 86(25): 5835-5838.