

基于预期流优化的空间网络引力模型^{*}

钱江海 韩定定[†]

(华东师范大学信息科学技术学院通信工程系, 上海 200241)

(2008 年 9 月 29 日收到, 2008 年 11 月 6 日收到修改稿)

从优化预期流的角度提出了一个空间网络模型, 其构建目标是使得整个网络的预期流总和最大化. 通过改变表征网络对节点适应度和对地理因素依赖程度的两个参数 α 和 γ 的取值, 网络的拓扑可从异质的星形类网络逐渐演变成同质的类似道路网的网络. 利用所提出的引力模型仿真了中国城市航空网, 并将其统计特征与实际网络进行了比较. 最后, 讨论了预期流和实际流(网络中实际的边权)之间的关系.

关键词: 空间网络, 预期流, 引力, 中国城市航空网

PACC: 0590, 0560, 0175

1. 引 言

近几年来, 人们发现众多实际网络都具有小世界现象^[1]和无标度特征^[2], 为理解这种新奇的拓扑结构及其演化规律, 复杂网络模型的研究不断深入开展^[3-6]. 大部分先前的研究都着重于网络的拓扑结构, 然而现实生活中的诸多网络往往是嵌入在一个真实的空间当中, 即它们的节点都有着确切的地理位置, 而它们的连边也同时受到地理因素的制约. 通信网络、电力网络、河流、航空、铁路以及地铁网络等便是空间网络的典型实例^[7-14]. 对于这类网络, 地理因素往往对拓扑及其演化起到重要影响. 这主要表现在空间网络中的节点总是趋向于和自己地理位置较近的节点连接^[15], 因为构建长距离的连边通常需要较大的花费和代价. 尽管在空间网络中地理限制对其拓扑影响巨大, 但研究表明它并不是唯一支配网络演化的机理. 事实上在一些空间网络中, 确实存在着少量的长距离连边, 并且这些连边大多与网络的轴心节点相连^[16]. 这意味着除了地理因素, 拓扑上的偏好依附同样影响着网络的演化^[2].

当实现拓扑上的偏好依附机理时, 一个可行的方法是考虑节点的适应度^[17, 18]. 所谓的适应度, 是指节点自身具有的某种内在属性, 它标志着节点在网

络中获取边的能力的大小. 对于不同类型的实际网络, 它可以分别代表节点的容量、人口、活跃程度及社交能力等. 一般认为, 具有较大适应度的节点往往具有较大的度值^[17]. 在地理因素对网络形成产生的限制方面, 主要存在两种观点: 一种认为节点间连边的概率随距离呈幂率衰减, 另一种则是呈指数衰减^[19-21]. 值得一提的是, 有学者提出每个节点都有其自己的感知范围, 节点只能与其感知范围内的节点连边, 而不同的节点的感知范围是不一样的, 这个想法将节点适应度和地理因素巧妙地结合了起来^[22].

以上这些研究为空间网络的建模提供了一些思路, 不过它们的总体思想大多是基于优化构建网络所需的代价. 然而在一个实际网络的建设中, 除了代价, 人们还非常关心网络的预期效益. 比如说在航空网络中, 如果航空公司准备开通一个新的航班, 他们会对该航班的载客流量进行预测, 只有当流量带来的效益大于其投资时, 这条航班才会被开通. 由此可见, 这种对预期效益的估计在人造网络的构建中扮演重要的角色. 出于这一点的考虑, 本文提出了预期流的概念, 并基于它提出了一种新的空间网络模型.

本文第 2 节提出预期流的概念, 并指出可用引力公式对预期流进行测度; 第 3 节提出了空间网络的引力模型; 第 4 节利用模型仿真实际的中国城市航空网, 并将其统计特性与实际网络相比较; 最后,

^{*} 上海科学技术发展基金(批准号: 06JC14082, 05XD14021)资助的课题.

[†] 通讯联系人. E-mail: ddhan@ee.ecnu.edu.cn

讨论了预期流与网络中实际流之间的关系。

2. 引力公式和预期流

传统的引力公式只是用来测度具有一定质量的两个物体之间的吸引力,而在此基础上改进的公式已在诸如地理、经济、贸易等领域得到了广泛的应用。地理学科的学者利用它来研究城市间的相互作用,或利用它来估计航空客流。在经济贸易领域里,它也同样被用来计算各种贸易流^[23-26]。改进的引力公式可表述如下:

$$T_{ij} = K \frac{M_i^\alpha M_j^\alpha}{D_{ij}^\gamma}, \quad (1)$$

其中 K 是一个常数, M_i 和 M_j 分别代表节点 i 和节点 j 的适应度, D_{ij} 一般被定义为欧氏距离,但也可以代表其他物理量,比如说时间。 α 和 γ 是两个参数,它们的值分别取决于网络对节点适应度和对地理因素的依赖程度,在描述城市间的相互作用时,可以从理论上证明 $\alpha = 1$,而 γ 的值已被诸多实际数据证实,在 0.2 至 2.7 之间波动^[23-25, 27]。最近的文献指出,韩国高速公路复杂系统中城市间的流量服从 $\alpha = 1, \gamma = 2$ 的引力公式^[28]。下面,我们通过一种逻辑上可信的方法来说明引力模型可用来测度节点之间的流量。大量实证研究表明,在一些网络中,比如说航空网络^[10],节点间的流量服从 $w_{ij} \propto (k_i k_j)^\theta$,其中 w_{ij} 是节点 i 和节点 j 之间的流量(即边权值), k_i 和 k_j 则分别是节点 i 和节点 j 的度值。现在考虑节点的适应度 M ,一般来说,它与节点的度是正相关的,可定义为 $M \sim k^\lambda$,代入公式 $w_{ij} \propto (k_i k_j)^\theta$,我们得到 $w_{ij} \propto (M_i M_j)^\gamma$,其中 $\alpha = \theta \lambda$ 。此外,在空间网络中,由于长距离的旅行或传输通常会花费较多的时间和代价,所以,节点间的流量将随它们之间距离的增加而衰减^[15, 26]。若采用幂率衰减的方式,即有 $w_{ij} \propto \frac{1}{D_{ij}^\gamma}$,将其与 $w_{ij} \propto (M_i M_j)^\gamma$ 结合,我们得到节点间的流量服从 $w_{ij} = K \frac{M_i^\alpha M_j^\alpha}{D_{ij}^\gamma}$,而这正是(1)式的表述。

因此,上述改进的引力公式适用于测度节点之间的流量,这就提供一种预测节点间流量的方法。因为(1)式并不依赖于网络的具体拓扑,所以我们可以网络构建之前,通过计算节点对的 T_{ij} 来估计该节点对之间可能存在的流量大小。因此,我们称 T_{ij} 为

预期流,它描述的是节点间尚未连边时的一种潜在的动力学。不过必须指出,预期流的值与网络中的实际流量是有区别的。这里我们特别强调, T_{ij} 描述的仅仅是产生于节点 i 和节点 j 之间的流量,即起源于节点 i (或者节点 j) 而终止于另一端的流量,它并不包括那些由其他节点对产生的,并经由某个通过 (i, j) 边的非直达路径传播的流量。这一点与真实网络中的流量是不同的。有关预期流和实际流的关系将在文章最后一部分予以讨论。

3. 空间网络的引力模型

如引言中所述,航空网络中的航班是否开通很大程度上取决于该航班的预期流。这一点对于那些受地理制约更为强烈的网络,比如说高速公路网,也是同样适用的。尽管在这类网络中代价对网络结构的影响要比其在航空网络中大得多,但预期效益依然是一个受关注的因素。事实上,即使构建一条连边的代价很小,但如果该边上没有流量,这样的连边也是毫无意义的,毕竟我们构建网络的目的是要服务其上的动力学。从这个意义上来说,预期流要比代价更重要。所以,在接下来要提出的模型里,我们关心的是如何最优整个网络的预期流,而代价则被简化为网络的边数。

现在我们考虑在一个二维方格里分布着一组节点,其地理信息和节点适应度已知。利用(1)式可以计算任意两个节点间的预期流 T_{ij} 。然后,优先连接那些预期流较大的节点对。详细地说,从 T_{ij} 最大的一对节点开始,按 T_{ij} 递减的顺序,依次连接相应的节点对,直到网络的边数(即构建网络的代价)达到预先设定的某个值为止。以上过程可描述为一个优化模型

$$\max W_{\text{exp}} = \sum_{i < j} T_{ij} \eta_{ij} = \sum_{i < j} K \frac{M_i^\alpha M_j^\alpha}{D_{ij}^\gamma} \eta_{ij}, \quad (2)$$

$$\text{s.t. } \sum_{i < j} \eta_{ij} = \epsilon, \quad (3)$$

这里 η_{ij} 是邻接矩阵元素, ϵ 是预先设定的网络的边数, W_{exp} 是网络的总预期流。上述的方法有可能造成某些无法连入网络的孤立节点,所以还需额外增加两个限制条件:

$$\begin{aligned} \sum_i \eta_{ij} &\geq 1, \quad j = 1, 2, 3, \dots, n, \\ \sum_j \eta_{ij} &\geq 1, \quad i = 1, 2, 3, \dots, n. \end{aligned} \quad (4)$$

(4)式保证了每个节点的度值至少为 1. 我们的模型有着实际含义,它表明在有限的成本约束下,人们是优先投资那些更迫切需要被建设的通路,这样不仅满足了系统最主要的需求,也能使运营商提高自己的效益.

利用上述模型,我们首先仿真了一个具有 39 个节点的小网络,目的在于直观地了解参数 α 和 γ 的变化对网络拓扑的影响.设定 $\epsilon = 39, K = 1$ (事实上 K 的取值对模型没有影响).通过改变 α 和 γ 的取值,我们得到了 4 个不同拓扑结构的网络,如图 1 所示.

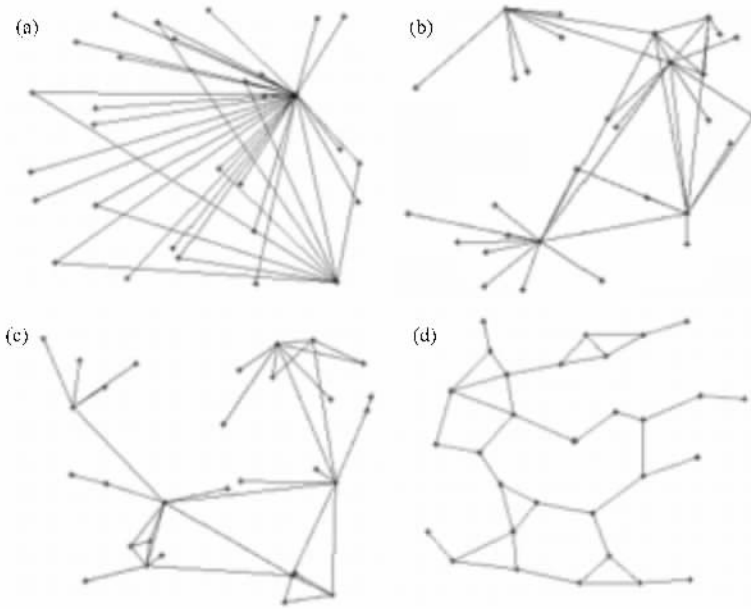


图 1 α 和 γ 取 4 类不同值形成的不同拓扑结构 (a) α 任意, $\gamma = 0$; (b) $\alpha = 1, \gamma = 1$; (c) $\alpha = 1, \gamma = 2$; (d) $\alpha = 0, \gamma$ 任意

在图 1(a)中, $\gamma = 0$, 此时,网络完全取决于节点的适应度,适应度大的节点极容易吸引其他节点并连接,所以网络完全由两个轴心点所支配,形成一种类似星状结构的拓扑.当 $\alpha = 1, \gamma = 1$ 和 $\alpha = 1, \gamma = 2$ 时(图 1(b)和(c)),网络呈现类似航空网络的拓扑结构.仔细观察可以看出,尽管两张图都存在轴心节点,但图 1(c)的拓扑比图 1(b)更趋于同质,这是由于地理制约的不断加强使得节点更趋向于和其附近的节点相连,从而网络的轴心效应也被进一步削弱.当 $\alpha = 0$ (图 1(d)),网络完全受限于地理因素并形成了类似道路网的结构特征.

的适应度 M 代表城市的人口,地理距离 D_{ij} 被定义为节点 i 和节点 j 之间的欧氏距离.设定 $\alpha = 1, \gamma = 1.5^{[25, 27]}$.

4. 中国城市航空网的仿真

为验证基于预期流的观点下的引力模型是否能够重建实际网络的统计特征,我们利用模型对中国城市航空网进行了模拟仿真,并将其统计特性与实际的网络^[4]作比较.其中节点代表城市而边代表航班,设定网络的节点数 $n = 121$,总边数 $\epsilon = 689$.节点

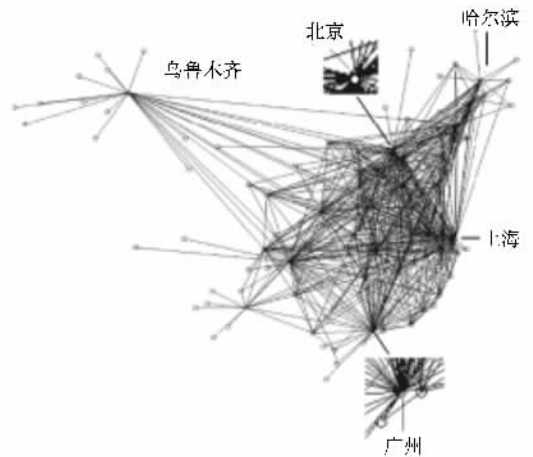


图 2 中国城市航空网络的仿真拓扑图

图 2 是仿真网络的拓扑结构图.其中北京,上海,广州,哈尔滨,乌鲁木齐是明显的轴心节点.这与实际网络是符合的.在真实的中国城市航空网里,度

值最大的三个城市是北京、上海和广州，而在我们的仿真中，它们则分别是上海、北京和武汉（广州位于第四）。造成这种差异的原因可能是用人口作为城市节点的适应度，虽然直观但未必准确，因为经济和行政因素同样是标志城市等级的重要指标。尽管如此，我们依然成功地重建了实际网络中存在的各轴心及其轴辐轮现象。

我们计算了网络的平均最短路径 $L^{[1]}$ 和 Pearson 相关系数 $r^{[29]}$ 。平均最短路径是网络中任意两节点间最短路径长度的平均值。Pearson 相关系数则描述了网络是同配还是异配的。模拟仿真表明网络的 $L_{sim} = 2.302$ ， $r_{sim} = -0.401$ ，这与实际值 $L_{real} = 2.263$ ， $r_{real} = -0.408^{[14]}$ 是很接近的。

图 3 显示了仿真网络的簇度相关性^[30]，它是指度值为 k 的节点的聚类系数的平均值 $C(k)$ 与 k 之间的关系。数据拟合显示，相关性满足 $C(k) \sim k^{-1}$ ，意味着仿真网络展现出与实际网络相同的层次性组织结构。

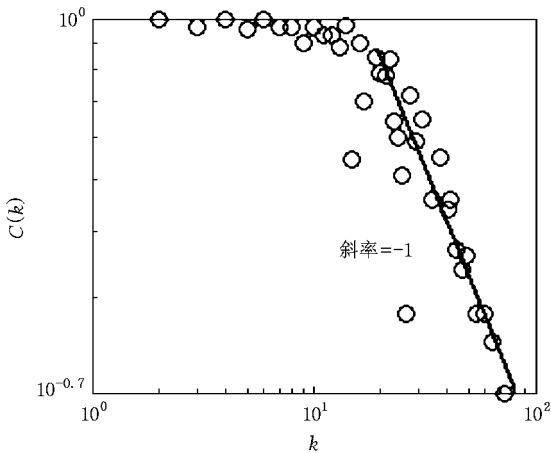


图 3 网络的簇度相关性

图 4 给出了网络的 Zipf 度分布，服从双段幂率分布，其中第一段的斜率指数 $\gamma_1 = -0.46$ 而第二段则是 $\gamma_2 = -2.3$ ， $k = 22$ 处为两段幂率转折点。实际网络中，这三个参数分别为 $\gamma_1 = -0.53$ ， $\gamma_2 = -2.05$ ， $k = 20^{[14]}$ 。有意思的是，在我们的模型里并没有引入所谓的“偏好依附”机理，但网络却自然地展现了与实际相符合的双段幂率特性。

5. 预期流与实际流

在第 2 节我们提出了预期流的概念，并用其预

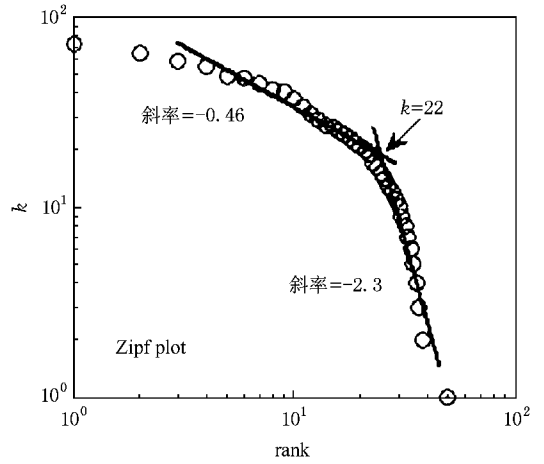


图 4 仿真网络的 Zipf 度分布

测节点间可能存在的流量大小。但我们同时也强调了预期流和实际流是有区别的。在这一部分，我们将讨论造成这一区别的原因以及预期流和实际流之间的关系。

我们认为，一种所谓的“转发机理”是造成预期流与实际流不同的根本原因。转发机理普遍存在于各种技术网络中，最为常见的便是 Internet。在 Internet 中，作为节点的路由器并不仅仅发送（或接收）起源于它（或终止于它）的流量，还大量地转发其他节点对的流。这些被路由器转发的流量加载于节点之上，并与预期流一起成为实际的流。同样，在航空网络中，城市间的一条航班并非都是起源且终止于这两个节点的客流，还有一些通过非直达航班转机的客流，但是这两类不同性质的客流是难以从实际数据中分离提取的。

接下来，我们将计算这两者的关系。我们遵循文献^[31]的假设，即认为两节点间的流通过最短路径传输，当有多条这样的路径时，该流被平均分配于这些路径中。因而，某个节点的实际点权 s_i 和预期流 T_{ij} 之间应满足

$$s_i = \sum_j T_{ij} + \sum_{\substack{(w, w') \\ w, w' \neq i}} T_{ww'} \frac{\sigma_{ww'}'(i)}{\sigma_{ww'}} \\ = \sum_{j \in \mathcal{A}(i)} T_{ij} + \left(\sum_{j \notin \mathcal{A}(i)} T_{ij} + \sum_{\substack{(w, w') \\ w, w' \neq i}} T_{ww'} \frac{\sigma_{ww'}'(i)}{\sigma_{ww'}} \right) \quad (5)$$

其中 $\sigma_{ww'}$ 是节点 w 与节点 w' 间的最短路径数，而 $\sigma_{ww'}'(i)$ 则是节点 w 与节点 w' 间通过节点 i 的最短路径数。 $\sum_{j \in \mathcal{A}(i)} T_{ij}$ 表示的是节点 i 与其邻接节点间的

流量之和, $\sum_{j \in i(i)} T_{ij}$ 则是节点 i 与非邻接节点间的流量之和. $\sum_{\substack{(w, w') \\ w, w' \neq i}} T_{ww'} \frac{\sigma_{ww'}(i)}{\sigma_{ww'}}$ 代表了节点 i 转发其他节点对的流量大小. 可以看出 (5) 式中括号内的项是与转发机理直接相关的, 而括号外的则与转发机理无直接关系. 所以对于像科研合作网这类无转发机理的社会网络, 应只具有第一项 $\sum_{j \in i(i)} T_{ij}$, 而对于存在转发机理的技术网络 (5) 式里所有的项都起作用. 读者可自行根据类似的方法定义实际边权, 这里不再赘述.

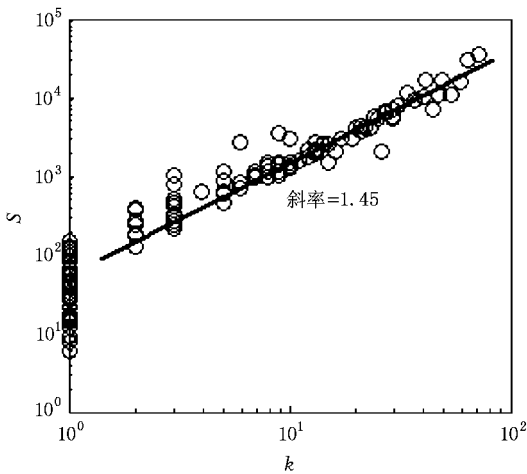


图 5 仿真网络的权度的相关性

根据 (5) 式, 我们计算了网络的权度相关性, 如图 5 所示. 仿真表明网络权度的相关性服从指数为 1.45 的非线性的幂率特征, 即 $s \sim k^{1.45}$, 这与实际航空网络中 $s \sim k^{1.37}$ [14] 基本相符.

6. 结 论

本文针对先前研究中忽略网络构建的预期效益这一点, 提出预期流的概念并用引力公式对其进行测度. 在此基础上, 我们提出了一个新的空间网络模型. 模型的目标是使整个网络的预期流最大化. 我们的模型有着实际的意义, 它表明在有限的成本约束下, 人们是优先投资那些更迫切需要被建设的通路, 这样不仅满足了系统最主要的需求, 也能使运营商提高自己的效益. 模型的拓扑可由两个参数 α 和 γ 来控制, 通过变换参数, 模型的拓扑可由类似星状的网络逐渐演变成分散同质的道路类网络. 不仅如此, 模型的统计特征也与实际网络相符合.

此外, 我们指出造成预期流与实际流不同的原因是网络中的转发机理, 而它们之间的关系可由 (5) 式给出. 我们同时指出对于无转发机理的社会网络来说, 其点强度仅仅包含该公式的第一部分. 这或许可进一步用来研究为什么社会网络的权度相关性呈现线性, 而技术网络中大多是非线性. 因为 (5) 式中括号内的非线性项很可能导致点权的增长率大大增加.

我们的模型不仅能帮助人们理解网络的形成, 还为网络上资源的优化分配提供思路. 因为模型中, 只要节点的适应度及其地理信息已知, 我们便可在实际网络形成之前对它的各种统计参量(如点权、边权、节点度)进行预测和计算, 这样我们可事先估计出网络的运行状况, 并将诸如带宽、容量等资源有目的地分配到必要的地方.

[1] Watts D J, Strogatz S H 1998 *Nature* **393** 440
 [2] Barabási A L, Albert R 1999 *Science* **286** 509
 [3] Guo J L, Wang L N 2007 *Acta Phys. Sin.* **56** 5635 (in Chinese)
 [郭进利、汪莉娜 2007 物理学报 **56** 5635]
 [4] Wu Y, Xiao J H, Wu Z Y, Yang J Z, Ma B H 2007 *Acta Phys. Sin.* **56** 2037 (in Chinese) [吴 晔、肖井华、吴智远 杨俊忠、马宝华 2007 物理学报 **56** 2037]
 [5] Barrat A, Barthélemy M, Vespignani A 2004 *Phys. Rev. Lett.* **92** 228701
 [6] Wang W X, Wang B H, Hu B, Yan G, Qu Q 2005 *Phys. Rev. Lett.* **94** 188702

[7] Latora V, Marchiori M 2005 *Phys. Rev. E* **71** 015103
 [8] Albert R, Albert I, Nakarado G L 2004 *Phys. Rev. E* **69** 025103
 [9] Guimer R, Mossa S, Turttschi A, Amaral L A N 2005 *Proc. Natl. Acad. Sci. USA* **102** 7794
 Guimer R, Amaral L A N 2004 *Eur. Phys. J. B* **38** 381
 [10] Barrat A, Barthélemy M, Vespignani A 2004 *Proc. Natl. Acad. Sci. USA* **101** 3747
 [11] Smith D A, Timberlake M 1995 *Urban Stud* **32** 287
 [12] Latora V, Marchiori M 2002 *Physica A* **314** 109
 [13] Han D D, Liu J G, Ma Y G, Cai X Z, Shen W Q 2004 *Chin. Phys. Lett.* **21** 1855

- Han D D , Liu J G , Ma Y G 2008 *Chin. Phys. Lett.* **25** 765
- Han D D , Qian J H , Liu J G 2009 *Physica A* **388** 71
- [14] Liu H K , Zhou T 2007 *Acta Phys. Sin.* **56** 106 (in Chinese)[刘宏鲲、周 涛 2007 物理学报 **56** 106]
- [15] Gastner M T , Newman M E J 2006 *Eur. Phys. J. B* **49** 247
- [16] Barthélemy M 2003 *Europhys. Lett.* **63** 915
- [17] Bianconi G , Barabási A L 2001 *Europhys. Lett.* **54** 436
- Bianconi G , Barabási A L 2001 *Phys. Rev. Lett.* **86** 5632
- [18] Naoki M , Hiroyoshi M , Norio K 2005 *Phys. Rev. E* **71** 036108
- [19] Kaiser M , Hilgetag C C 2004 *Phys. Rev. E* **69** 036103
- [20] Xie Y B , Zhou T , Bai W J , Chen G R , Xiao W K , Wang B H 2007 *Phys. Rev. E* **75** 036106
- [21] Yook S H , Jeong H , Barabási A L 2002 *Proc. Natl. Acad. Sci. USA* **99** 13382
- [22] Xulvi-Brunet R , Sokolov I M 2007 *Phys. Rev. E* **75** 046117
- [23] Zipf G K 1947 *Am. Sociol. Rev.* **12** 677
- [24] Long W H 1970 *Ann. Regional. Sci.* **4** 97
- Howrey E P 1969 *J. Regional Sci.* **9** 215
- [25] Fotheringham A S 1981 *Ann. Assoc. Am. Geographers* **71** 425
- [26] Jin F J 2001 *Geographical Research* **20** 31 (in Chinese)[金凤君 2001 地理研究 **20** 31]
- [27] Liu J S , Chen Y G 2000 *Scientia Geographica Sinica* **20** 6 (in Chinese)[刘继生、陈彦光 2000 地理科学 **20** 6]
- [28] Jung W S , Wang F Z , Stanley H E 2008 *Europhys Lett.* **81** 48005
- [29] Newman M E J 2003 *Phys. Rev. E* **64** 026126
- [30] Ravasz E , Barabási A L 2003 *Phys. Rev. E* **67** 026112
- [31] Goh K I , Kahng B , Kim D 2001 *Phys. Rev. Lett.* **87** 278701

Gravity model for spatial network based on optimal expected traffic^{*}

Qian Jiang-Hai Han Ding-Ding[†]

(Department of Communication Engineering , School of Information Science and Technology , East China Normal University , Shanghai 200241 , China)

(Received 29 September 2008 ; revised manuscript received 6 November 2008)

Abstract

In this paper , we propose a spatial network model based on the optimal expected traffic . The model maximizes the total expected traffic of the network . By changing two parameters α and γ which represent the dependence of the network on the fitness and geographical constraints respectively , the model can vary its topology from heterogeneous star-like networks to homogeneous road-like networks . We use our model to simulate the Chinese city airline network and compare it with the real data . In the end of the paper , we discuss the relationship of the expected and the real traffic (namely the real weight on the network) .

Keywords : spatial network , expected traffic , gravity , Chinese city airline network

PACC : 0590 , 0560 , 0175

^{*} Project supported by the Shanghai Development Foundation for Science and Technology (Grant Nos. 06JC14082 and 05XD14021).

[†] Corresponding author. E-mail : ddhan@ee.ecnu.edu.cn