

二维有限能量约束下最优导航问题的理论分析*

黎勇 钊斐玲 樊瑛 狄增如†

(北京师范大学管理学院系统科学系, 北京师范大学复杂性研究中心, 北京 100875)

(2011年11月10日收到; 2012年6月3日收到修改稿)

最近, Li 等研究了在 Kleinberg 导航模型中引入总能量 $A = cN$ 约束下的最优导航问题, 其中 A 为网络中所有长程连边的长度之和, c 为正常数, N 为网络节点总数. 他们通过在 1 维和 2 维导航模型中的模拟结果推测, 在有限能量约束下 Kleinberg 导航模型中按照幂律方式添加长程连边的最优幂指数应该是 $\alpha = d + 1$, 其中 d 为导航模型的维数. 本文在平均场理论下, 建立了 2 维有限能量约束下的导航过程的动态微分方程, 通过对该方程进行数学分析以及数值求解, 从理论上证明了当网络规模足够大且总能量相对较小时, 2 维有限能量约束下的最优导航幂指数确实为 $\alpha = 3$, 这一结果证实了 Li 等之前的推测.

关键词: 网络导航, 能量约束, 复杂网络, 空间网络

PACS: 89.75. Hc, 89.75.Fb

1 引言

近几十年来, 科学家们掀起了一股研究复杂网络的热潮. 除了网络的小世界特性、无标度特性以及划分网络社团结构之外, 网络上的动力学行为也是复杂网络中的一个研究热点^[1,2]. 其中, 网络可导航问题这一动力学行为一直以来就受到了广泛关注^[3-6].

早在 20 世纪 60 年代, Milgram 等设计了一系列著名的社会搜寻试验用以研究社会网络中的朋友关系网络的信息传递能力^[7,8]. 在 Milgram 的实验中, 随机选取的 296 位试验者被要求通过他们认识的朋友将信件传递给指定目标. 最终统计发现, 共计 64 封信件被成功传递到了指定的目标, 并且这些成功实验所需的传递次数平均为 6 次. 该实验结果就是后来广为流传的“六度分离”现象. 从此人们开始意识到社会网络中任何个体都可能只需要通过较少的几次搜索就能够被发现. 后来, Dodds 等又在全球范围内重复了 Milgram 的搜寻试验, 并且得到了类似结论, 他们发现大部分搜索仅需要经过大约 5 到 7 步就能够完成^[9].

当“六度分离”现象被广为人知后, 科学家们就试图对其进行理论研究和解释. 其中, 一个想法就是通过网络模型来抽象 Milgram 实验中的朋友关系网络, 希望能够在该网络上重现“六度分离”这一结果. 仔细分析 Milgram 的实验, 可以发现“六度分离”现象的产生需要满足两个基本条件. 第一, 朋友关系网络中的任意两个个体之间要存在大量的较短连通路程; 第二, 处于该网络中的个体在一定的搜索策略下能够有效地找到通往目标个体的较短路径. 也就是说, 能够重现“六度分离”现象的任何一个网络模型首先应该要满足网络中的任何节点之间存在着大量的较短连通路程. 其次, 考虑到当网络规模较大时处于网络中的个体通常不了解网络的全局结构而仅仅具备有关网络结构的局部信息这一事实, 该网络模型还应该具备较好的导航能力, 即处于该网络中的个体凭借对网络整体结构的一些局部信息就能够有效地找出通往目标节点的较短路径.

迄今为止, 能够最好地解释“六度分离”现象的复杂网络模型便是 Kleinberg 在 2000 年提出的网络可导航模型^[10,11]. 在其模型中, Kleinberg 使用一个规模为 $n \times n$ 的 2 维规则网格来抽象真实的朋友

* 中央高校基本科研业务费专项资金、国家自然科学基金(批准号: 60974084, 61174150)和 NCET-09-0228 资助的课题.

† E-mail: zdi@bnu.edu.cn

关系网络. 网络中的每一个节点代表一个个体, 而节点之间的连边则代表个体之间存在的友谊关系. 模型假定每个节点都与其网格距离 (曼哈顿距离) 为 p 的所有节点之间存在双向的短程连接, 同时该节点还拥有 q 条单向长程连边. 进一步, 网络中的任意两个节点 u 和 v 之间存在一条有向长程连边的概率服从下式所示的幂律分布:

$$P_K(u, v) = \frac{d(u, v)^{-\alpha}}{\sum_{w \neq u} d(u, w)^{-\alpha}}, \quad (1)$$

其中, $d(u, v)$ 是节点 u 与 v 之间的网格距离. 显然, 当参数 α 不断变化的时候, 网络的空间结构也会随之变化. 另外, Kleinberg 还提出了一种基于局部信息的分散式贪婪算法. 由于该算法仅需要每一个信息持有者具备网络结构的局部信息, 即当前的信息持有者仅仅知道目标节点以及与其存在连边 (朋友关系) 的所有邻居节点在网络中的具体坐标位置, 因而使用该算法来描述 Milgram 试验中的搜索策略就比较符合实际. 具体来讲, 该算法要求网络中当前的信息持有者在每一次的传递过程中都将信息传递给离目标节点网格距离最近的邻居节点.

在前述模型框架下, Kleinberg 证明了在 2 维网格中, 对于任意给定的 p 和 q , 使得网络模型具备最优导航能力 (通过最少的搜索次数找到目标节点) 的幂指数为 $\alpha = 2$, 并且相应的搜索时间复杂度上限为 $O(\log^2 n)$ [10]. 该结论可以被推广到任意的 d 维 (整数维) 网格中, 即当且仅当幂指数满足 $\alpha = d$ 时, 分散式搜索的时间复杂度最低, 即网络的导航能力最强. 在此基础上, Roberson 等对分形维度下的 Kleinberg 导航模型进行了讨论 [12]. 他们在 Sierpinski 地毯上进行模拟并且证明了分形网络中的最优导航幂指数也为 $\alpha = d_f$, 其中 d_f 为分形网络的维度. 也就是说通过在规则网格上添加长程连边的方式构建起来的导航模型中, 具备最优导航能力的网络模型都应该满足幂指数 $\alpha = d$, 其中 d 为该规则网络的维度.

后来, Martel 等在研究 Kleinberg 的导航模型时发现最优幂指数下的导航时间复杂度为紧约束即 $O(\log^2 n)$, 并发展了一套相应的分析空间网络上导航问题的数学方法 [13]. 2009 年, Carmi 等使用主方程的方法进一步分析了 Kleinberg 导航模型的渐进行为, 给出了导航时间复杂度和幂指数 α 之间更

为精确的函数关系 [14]. 几乎同时, Caretta Cartozo 等也使用随机马尔科夫过程的方法得到了与 Carmi 等相同的结论 [15].

至此为止, 已有的导航模型都是从 Kleinberg 的导航模型出发, 假定网络中的每一个节点都拥有至少一条长程连边, 也就是说网络中所有节点长程连边的长度总和并不固定, 会随着幂指数 α 的变化而变化. 事实上, 在许多现实的网络中, 维系或者创建一条长程连边都可能会产生巨大的成本. 以改建一个现存的地铁网络为例, 创建一条新的长程连边意味着新建一条地铁线路, 需要付出高额的成本, 而通常财务预算却是有限并且事先给定的. 因此, 一个更加合理的假设是网络中的所有长程连边长度总和应该受到一定的限制. 基于此想法, 杨华等首先在他们的空间网络模型中引入了总能量 (即长程连边长度总和) 约束, 他们研究发现 1 维情况下使得网络平均距离最短的幂指数为 $\alpha = 2$ [16]. Li 等在 Kleinberg 导航模型中添加了总能量 (长程连边长度总和) $A = cN$ 的限制, 其中 c 为常数, N 为网络总节点数 [17]. 他们研究了 1 维和 2 维有限能量约束下的导航时间复杂度与幂指数 α 的关系. 模拟结果显示, 在 1 维情况下使得网络具有最优导航能力的最优幂指数为 $\alpha = 2$, 而 2 维情况下的最优幂指数则满足 $\alpha = 3$. 因此, 他们推断 Kleinberg 导航模型在有限能量约束下的最优导航幂指数应该满足 $\alpha = d + 1$, 其中 d 为相应的网络维数. Bianconi 等实证发现美国航空网的幂指数为 $\alpha = 3.0 \pm 0.2$, 该结果进一步支持了 Li 等关于 2 维有限能量约束下的导航最优幂指数为 $\alpha = 3$ 这一推断 [18].

然而, Li 等关于有限能量约束下的最优导航幂指数的结论主要来自于计算机模拟和一些简单的数学分析, 并没有给出严格的数学证明. 如果能够从理论上更加严格地证实 Li 等发现的结论将具有重要的理论意义. 在前一个工作中, 我们已经严格证明当网络规模足够大并且总能量相对较小时, 使得 1 维有限能量约束下的空间网络导航能力最强的最优幂指数为 $\alpha = 2$, 并且给出了其导航时间复杂度与幂指数 α 的精确解 [19]. 本文将对更具现实意义的 2 维有限能量约束下的 Kleinberg 导航问题进行理论分析, 给出 2 维导航过程的微分方程, 并试图通过数学分析及微分方程的数值解从理论上说明使得 2 维有限能量约束下的空间网络具有最优导航能力的幂指数为 $\alpha = 3$.

2 有限能量约束下 2 维导航过程的动态方程

与 Li 等的有限能量约束导航模型类似, 我们引入一个满足循环边界条件且网络规模为 $2n \times 2n$ 的 2 维网格, 网格上的每个节点都与其网格距离为 1 的节点之间存在双向的短程连边. 另外假定通过幂律方式添加的长程连边长度之和即总能量满足 $\Lambda = 4cn^2$, 其中 c 为给定正常数. 已知区间 $[m, M]$ 上服从幂指数为 r 的幂律分布的概率密度函数满足

$$P(l) = k(r, m, M) \cdot l^{-r}, \quad l \in [m, M], \quad (2)$$

其中, $k(r, m, M) = 1 / \int_m^M l^{-r} dl$ 为归一化因子. 本文按照如下方式为网络中任意节点 u 添加有向长程连边: 首先按照 (2) 式随机产生一个服从幂律分布的网格距离 l , 然后从离节点 u 网格距离为 l 的所有节点中随机选取一个节点 v 建立一条从 u 到 v 的有向长程连边, 同时相应地将剩余总能量减少 l . 按照这种方式, 每次从 2 维网格中随机挑选一个节点产生一条有向的长程连边, 直到总能量 Λ 耗尽, 最终可以得到总能量受到约束的 2 维导航模型.

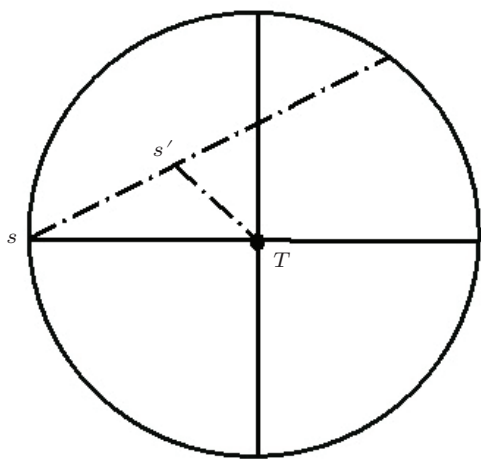


图 1 导航过程示意图

我们将首先建立在上述 2 维有限能量约束导航模型中使用 Kleinberg 提出的分散式贪婪算法进行导航的动态方程. 导航过程示意图如图 1 所示: 其中 s 为信息当前所处位置, T 为导航的目标节点, 假定 s' 为经过一次导航后信息所处位置, 令 $|ss'| = l$, 直线 ss' 与 sT 之间的夹角为 θ , 并且用 $s(t)$ 表示被传递的信息经过时间 t 之后累计向目标节点 T 前进的距离, 因此一定有 $s(0) = 0$ 成立.

另外, 我们假定初始时刻出发节点与目标节点 T 之间的网格距离为 n . 当信息在 t 时刻从当前节点 s 被传递到节点 s' 时相当于信息向目标节点 T 前进了如下距离:

$$\Delta s = (n - s) - \sqrt{(n - s)^2 + l^2 - 2l(n - s) \cos \theta}, \quad (3)$$

为了书写简便, 此处我们用 s 代替 $s(t)$.

由前文可知, 2 维空间网络上的长程连边的长度满足幂指数为 r 的幂律分布. 由定义可以知道二维网格上的节点 (x_1, y_1) 与 (x_2, y_2) 之间的网格距离满足 $d = |x_1 - x_2| + |y_1 - y_2|$, 其中 x_1, x_2 表示节点的横坐标, y_1, y_2 表示节点的纵坐标. 也就是说 2 维上的网格距离可以表示为水平方向和垂直方向上的两个投影之和. 不难证明, 当网络规模足够大时, 将 2 维网格上服从幂律分布的网格距离投影到 1 维之后的得到的新的网格距离同样服从幂律分布且保持幂指数不变, 详细证明过程请参考文献 [20]. 进一步, 我们可以知道选择向 ss' 方向进行信息传递时向目标节点前进的期望距离满足

$$E_s = \int_1^{2(n-s) \cos \theta} k(r, 1, 2(n-s) \cos \theta) \times l^{-r} \Delta s dl. \quad (4)$$

显然, 在 s 处可以选择导航的方向 θ 是服从区间 $[-\pi/2, \pi/2]$ 上的均匀分布. 因此, 信息在当前节点 s 通过长程连边向目标节点 T 前进的期望距离可以表示为

$$E_{Js} = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} E_s d\theta. \quad (5)$$

由于整个 2 维网络存在总能量 $\Lambda = 4cn^2$ 的约束条件, 因此整个网络中的长程连边数量也将会受到限制. 不难知道按照 (2) 式的方式添加的长程连边期望长度满足

$$E(L_r) = \frac{\int_1^n l^{1-r} dl}{\int_1^n l^{-r} dl}. \quad (6)$$

也就是说在总能量 Λ 的约束下, 整个网络中长程连边的平均数量为

$$E(N_r) = \frac{\Lambda}{E(L_r)}. \quad (7)$$

又因为网络中的所有节点都是同质的, 所以单个节点上的长程连边数量近似服从参数为 $\lambda = \frac{E(N_r)}{4n^2}$ 的泊松分布. 进而, 任意节点上存在一条长程连边的概率可以表示为 $\lambda e^{-\lambda}$. 当网络

规模足够大且总能量约束相对较小, 即 n 足够大且 c 相对较小时, 单个节点上存在 2 条及以上数量的长程连边的概率可以忽略不计. 为了进一步简化我们的数学分析过程, 本文只讨论 c 相对较小的情况. 在此前提下, 我们可以得到在总能量 Λ 的约束下, 信息从 s 处经过一次传递过程 (长程连边或者短程连边) 向目标节点 T 前进的期望距离满足

$$E(J_s) = \lambda e^{-\lambda} \cdot E_{J_s} + \left[1 - \lambda e^{-\lambda} \cdot \int_{-\pi/2}^{\pi/2} \frac{1}{\pi} d\theta \times \int_1^{2(n-s)\cos\theta} k(r, 1, 2(n-s)\cos\theta) \times l^{-r} dl \right]. \quad (8)$$

最终, 在一系列的简化过程之后, 2 维有限能量约束下的导航过程可以抽象为如下的一个微分方程:

$$\begin{aligned} \frac{ds(t)}{dt} &= E(J_s), \\ s(0) &= 0, \end{aligned} \quad (9)$$

其中 $s(t)$ 表示截至 t 时刻信息向目标节点 T 累计前进的距离.

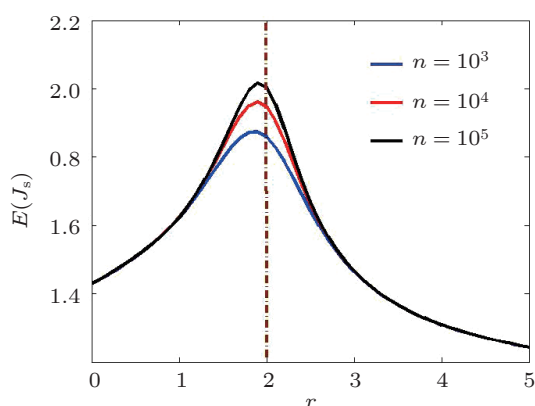


图 2 $E(J_s)$ 的网络规模效应

考虑到精确求解方程 (9) 的复杂性, 我们首先通过分析其右端项即 (8) 式的单调性来获取一些定性分析结果, 从而判断 2 维有限能量约束下导航过程的最优幂指数. 通过对右端项的数值模拟, 我们得到了 $E(J_s)$ 在不同的网络规模 n 的情况下随着幂指数 r 的变化情况. 图 2 给出了 n 不同取值情况下 $E(J_s)$ 在 $s(0) = 0$ 时刻, 即信息从出发节点经过一次传递后向目标节点 T 前进的距离随着幂指数 r 的变化情况. 其中 n 的取值从下往上依次为 $10^3, 10^4$ 以及 10^5 , 并且保持参数 $c = 1$. 从图 2 可以看出随着网络规模 n 逐渐增加, 使得 $E(J_s)$ 取

最大值的 r 将逐渐向 2 靠近. 可以合理推测当 n 趋于无穷时, $E(J_s)$ 将可能会在 $r = 2$ 处取得最大值. 为了验证我们的推测, 我们进一步给出了在参数 $c = 1$ 时的微分方程 (9) 的数值解, 具体结果如图 3 所示. 图 3 的纵坐标为 T/n , 其中 T 表示整个导航过程需要的时间, 而 n 表示出发节点到目标节点之间的网格距离. 从图 3 可以看出有限能量约束下的导航过程的时间复杂度与整个导航过程前进的网格距离 n 之间呈线性关系, 该结论与之前“不存在总能量约束的 Kleinberg 导航模型中最优导航时间复杂度为 $\log n$ 的多项式”这一结论有显著区别. 另外, 从图 3 中的数值结果我们也可以看出, 使得网络具备最优导航能力的幂指数确实在 $r = 2$ 附近. 考虑到在建立动态方程的过程中, 我们所进行的一些简化处理以及进行计算机模拟过程中不可避免的误差, 图中的最优幂指数与 2 存在一定的偏离是合理的. 另外, 仔细观察图 2 和图 3 我们也可以发现, 随着网络规模不断增加, 最优幂指数也逐渐向 $r = 2$ 靠近, 但是靠近速度相对较慢, 这一现象也可以从我们之前关于 1 维有限能量约束下的最优导航问题的研究工作中得到验证.

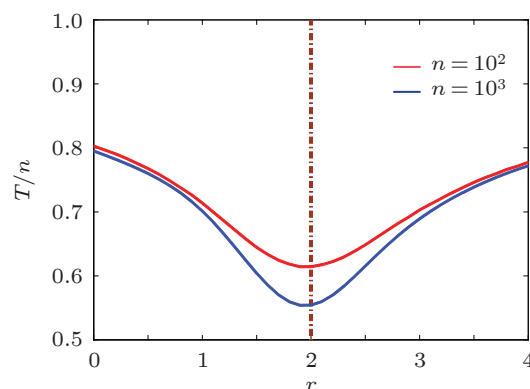


图 3 导航时间复杂度的数值解

我们知道, 在 2 维网格中, 与任意节点的网格距离为 l 的节点数量与 l 成正比. 也就是说在我们的 2 维导航模型中任意给定的网格距离为 $d(u, v)$ 的两个节点 u 和 v 之间存在长程连边的概率应该满足 $p(u, v) \propto d(u, v)^{-(r+1)}$. 于是, Li 等的导航模型中的幂指数 α 与本文的导航模型中的幂指数 r 的关系为 $\alpha = r + 1$. 经过以上数学分析, 我们证明了当网络规模足够大且总能量相对较小时, 使得 2 维有限能量约束下的空间网络具备最优导航能力的幂指数满足 $\alpha = 3$, 该结论与 Li 等通过模拟发现的结论能够很好地符合.

3 结论

长久以来,网络可导航问题一直是复杂网络研究中的一个热点问题,尤其是在 Kleinberg 的导航模型之后,更是引发了国内外学者的广泛研究热情.然而对于具备充分现实意义的有限能量约束下的可导航问题的研究才刚刚起步. Li 等率先在 Kleinberg 的导航模型中引入总能量约束的限制,通过在 2 维空间网络上进行模拟发现在 2 维情况下导航的最优幂指数应该为 $\alpha = 3$, 这一结论与不存在总能量约束下的最优导航幂指数为 $\alpha = 2$ 这一结论存在显著差异. Li 等对此结论进行了一些简单的

分析,然而并没有给出严格的数学证明. 本文在此基础上通过一系列的简化和抽象,建立起了 2 维有限能量约束下的导航过程的微分方程. 通过对该微分方程进行数学分析以及数值求解,我们从理论上说明了当网络规模足够大且总能量相对较小时,使得 2 维有限能量约束下的空间网络具备最优导航能力的幂指数确实满足 $\alpha = 3$. 该结论证实了 Li 等之前通过模拟结果做出的合理推断. 本文的工作不但能够从理论上丰富和完善空间网络上的导航问题研究,同时也能够为改善一些现实交通网络的运行效率提供启示,具有相当的理论意义和现实意义.

-
- [1] Li S B, Wu J J, Gao Z Y, Lin Y, Fu B B 2011 *Acta Phys. Sin.* **60** 050701 (in Chinese) [李树彬, 吴建军, 高自友, 林勇, 傅白白 2011 物理学报 **60** 050701]
 - [2] Xu D, Li X, Wang X F 2007 *Acta Phys. Sin.* **56** 1313 (in Chinese) [许丹, 李翔, 汪小帆 2007 物理学报 **56** 1313]
 - [3] Watts D J, Strogatz S H 1998 *Nature* **393** 6684
 - [4] Barabasi A L, Albert R 1999 *Science* **286** 509
 - [5] Girvan M, Newman M E J 2004 *Proc. Natl. Acad. Sci.* **99** 7821
 - [6] Du H F, Li S Z, Marcus W F, Yue Z S, Yang X S, 2007 *Acta Phys. Sin.* **56** 6886 (in Chinese) [杜海峰, 李树苗, Marcus W F, 悦中山, 杨绪松 2007 物理学报 **56** 6886]
 - [7] Milgram S 1967 *Psychology Today* **2** 60
 - [8] Travers J, Milgram S 1969 *Sociometry* **32** 425
 - [9] Dodds P S, Muhamad R, Watts D J 2003 *Science* **301** 827
 - [10] Kleinberg J 2000 *Nature* **406** 845
 - [11] Kleinberg J 2000 *Proceedings of the thirty-second annual ACM symposium on Theory of computing* 163–170
 - [12] Roberson M R, Ben-Avraham D 2006 *Phys. Rev. E* **74** 17101
 - [13] Martel C, Nguyen V 2004 *Proceedings of the Symposium on Principles of Distributed Computing*, ed. Kuttan, S. (ACM Press, New York) 179–188
 - [14] Carmi S, Carter S, Sun J, Ben-Avraham D 2009 *Phys. Rev. Lett.* **102** 238702
 - [15] Caretta Cartozo C, De Los Rios P 2009 *Phys. Rev. Lett.* **102** 238702
 - [16] Yang H, Nie Y C, Zeng A, Fan Y, Hu Y Q, Di Z R 2010 *EPL* **89** 5800
 - [17] Li G, Reis S D S, Moreira A A, Havlin S, Stanley H E, Andrade Jr. J S 2010 *Phys. Rev. Lett.* **104** 018701
 - [18] Bianconi G, Pin P, Marsilli M 2009 *Proc. Natl. Acad. Sci.* **106** 11433
 - [19] Li Y, Zhou D, Hu Y Q, Zhang J, Di Z R 2010 *EPL* **92** 58002
 - [20] Hu Y Q, Li Y, Di Z R, Fan Y 2010 arXiv: 1010.18

Theoretical analysis on optimal navigation with total energy restriction in a two-dimensional lattice*

Li Yong Dou Fei-Ling Fan Ying Di Zeng-Ru[†]

(Department of Systems Science, School of Management and Center for Complexity Research, Beijing Normal University, Beijing 100875, China)

(Received 10 November 2011; revised manuscript received 3 June 2012)

Abstract

Recently, a certain total energy constraint $A = cN$ was introduced into the Kleinberg's navigation model, where A is the total length of the long-range connections, c is a positive constant and N is the network size. The simulation results obtained in the one and two-dimensional cases indicate that with total cost restricted the optimal power-law exponent for adding extra long-range links between any two nodes seems to be $\alpha = d + 1$, where d is the dimension of the underlying lattice. Based on the mean field theory, the navigation process on the two-dimensional energy constrained navigation model is described by dynamical equations in this paper. Based on our theoretical analysis and the numerical results of the dynamical equations, we prove that for large networks and comparatively small total energy, the optimal power-law exponent is $\alpha = 3$ for the two-dimensional case. Our results can perfectly correspond to simulations reported previously.

Keywords: navigation, cost restriction, complex networks, spatial networks

PACS: 89.75. Hc, 89.75.Fb

* Project supported by the Fundamental Research Funds for the Central Universities and NSFC (Grants Nos. 60974084, 61174150) and NCET-09-0228.

[†] E-mail: zdi@bnu.edu.cn