

网络、通信、安全

PH-MaxFlow算法发现Web社区

郭希娟¹, 刘静²

燕山大学 信息科学与工程学院, 河北 秦皇岛 066004

收稿日期 2008-1-11 修回日期 2008-4-21 网络版发布日期 2009-2-9 接受日期

摘要 HITS是一种经典的利用链接技术计算网页权威值和中心值的算法,它能够快速发现主题相关网页,其缺点是会发生“主题偏移”现象,首先提出PHITS算法,在一定程度上抑制了这种现象的发生。运用该方法提取权威值高的页面,作为PH-MaxFlow算法的种子节点,使得发现的Web社区更精确。同时提出了一种有效的评价Web社区的标准,用这个标准对原始最大流算法和提出的PH-MaxFlow算法进行比较,从而得出PH-MaxFlow算法发现的Web社区与主题更相关。

关键词 [Web社区](#) [基于超链接分析的主题搜索算法](#) [最大流算法](#)

分类号

Mining Web communities with PH-MaxFlow algorithm

GUO Xi-juan¹, LIU Jing²

College of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China

Abstract

HITS is a classical algorithm for the computation of the authority value and hub value of Web pages using link technology, it can mine Web communities related to some topic quickly, but sometimes there is “Topic Drift” phenomenon. This paper presents PHITS algorithm controlling the “Topic Drift” phenomenon to a certain extent, the PH-MaxFlow algorithm using the pages with high authority value as seeds can mine precise communities. An effective method is presented to appraise the identified Web communities. The results of experiment show that the PH-MaxFlow algorithm can mine more reasonable Web communities.

Key words [Web communities](#) [Hyperlink-Induced Topic Search \(HITS\) algorithm](#) [maximum flow algorithm](#)

DOI: 10.3778/j.issn.1002-8331.2009.05.033

通讯作者 郭希娟 l6616j@yahoo.com.cn

扩展功能

本文信息

- ▶ [Supporting info](#)
- ▶ [PDF\(706KB\)](#)
- ▶ [\[HTML全文\]\(0KB\)](#)
- ▶ [参考文献](#)

服务与反馈

- ▶ [把本文推荐给朋友](#)
- ▶ [加入我的书架](#)
- ▶ [加入引用管理器](#)
- ▶ [复制索引](#)
- ▶ [Email Alert](#)
- ▶ [文章反馈](#)
- ▶ [浏览反馈信息](#)

相关信息

- ▶ 本刊中 [包含“Web社区”的相关文章](#)
- ▶ 本文作者相关文章
- [郭希娟](#)
- [刘静](#)