



Universiteit
Leiden
The Netherlands

Algorithms for Analyzing and Mining Real-World Graphs

Takes, F.W.

Citation

Takes, F. W. (2014, November 19). *Algorithms for Analyzing and Mining Real-World Graphs*. Retrieved from <https://hdl.handle.net/1887/29764>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/29764>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/29764> holds various files of this Leiden University dissertation.

Author: Takes, Frank Willem

Title: Algorithms for analyzing and mining real-world graphs

Issue Date: 2014-11-19

Stellingen

behorende bij het proefschrift

Algorithms for Analyzing and Mining Real-World Graphs

door Frank W. Takes

1. Het bepalen van de diameter van een sparse ongerichte samenhangende graaf met n knopen en m takken kan, in het geval dat deze graaf aan de “small-world” eigenschap voldoet, veel sneller dan binnen de traditionele APSP-tijdscomplexiteit van $O(mn)$ (Hoofdstuk 2).
2. Het bepalen van de eccentriciteitsdistributie en afgeleide extreme afstandsmaten van een sparse ongerichte samenhangende graaf met n knopen en m takken kan, door slim gebruik te maken van onder- en bovengrenzen, vaak veel sneller dan binnen de traditionele tijdscomplexiteit van $O(mn)$ (Hoofdstuk 3 en Hoofdstuk 4).
3. Bij het bepalen van de afstand tussen twee knopen in een samenhangende graaf gebruikmakend van zogenaamde “landmarks”, is het belangrijk om bij de selectie van deze landmarks naast de centraliteit ook een gespreide ligging van de landmark-knopen in acht te nemen (Hoofdstuk 5).
4. Bij het zoeken naar wie de “celebrities” binnen een online sociaal netwerk zijn, blijken twee factoren een belangrijke rol te spelen: de graad van een gebruiker en de genormaliseerde hoeveelheid driehoeksrelaties onder de vrienden van deze gebruiker (Hoofdstuk 6).
5. Bij het gericht navigeren door een gestructureerd informatiesysteem zoals Wikipedia, is het mogelijk om met een aantal lokale eigenschappen van de doelpagina te voorspellen hoe moeilijk het voor een gebruiker is om deze doelpagina te bereiken (Hoofdstuk 7).
6. Bij het gericht navigeren door een gestructureerd informatiesysteem zoals Wikipedia zijn gebruikers in staat om een nuttige deelverzameling van de pagina’s in het netwerk te selecteren die efficiënt navigeren mogelijk maakt (Hoofdstuk 8).
7. Bij data mining bestaat het risico dat het onderzoek neerkomt op het ondersteunen van andere disciplines op het gebied van data-analyse, in plaats van het leveren van een nuttige bijdrage aan de informatica.
8. Om een beter beeld van de herbruikbaarheid van een techniek te krijgen, dient in wetenschappelijke artikelen nadrukkelijker te worden vermeld welke algemene technieken en welke domein-specifieke technieken worden gepresenteerd.
9. Binnen het universitaire informatica-onderwijs kunnen online sociale netwerken worden ingezet als eenvoudig voorbeeld om de verschillende formele concepten van een graaf uit te leggen.
10. Het “filteren van data” is een multi-interpreteerbare activiteit.

Propositions

for the PhD thesis

Algorithms for Analyzing and Mining Real-World Graphs

by Frank W. Takes

1. Determining the diameter of a sparse undirected connected graph with n nodes and m links can, in case the graph's structure adheres to the "small-world"-property, be done much faster than within the traditional APSP time complexity of $O(mn)$ (Chapter 2).
2. Determining the eccentricity distribution and derived extreme distance measures of a sparse undirected connected graph with n nodes and m links can often be done much faster than within the traditional APSP time complexity of $O(mn)$ by making use of smart lower and upper bounds (Chapter 3 and Chapter 4).
3. When computing distances between nodes in a connected graph using the so-called "landmark" technique, it is important to select landmark nodes that are both central and well spread over the graph (Chapter 5).
4. When searching for "celebrity" users within an online social network, two factors seem to play a significant role: the degree of the user and the normalized number of triangular connections amongst the friends of a user (Chapter 6).
5. Considering a goal-oriented search within an information network such as Wikipedia, it is possible to use local properties of the goal page to determine the difficulty (for a human) of finding this goal page (Chapter 7).
6. Considering a goal-oriented search within an information network such as Wikipedia, users are able to select a useful subset of the pages of the network that makes efficient navigation possible (Chapter 8).
7. In data mining research, there is a risk of solely supporting other disciplines with data analysis, instead of giving a useful contribution to the field of computer science.
8. To get a better idea of the reusability of a technique, in scientific articles there should be a more clear indication of which techniques are general and which techniques are domain specific.
9. An academic computer science curriculum can benefit from the concept of an online social network, as it can be used as an easy example to explain the formal aspects of graphs.
10. "Filtering data" is multi-interpretable.