

# Personal Information Leakage through Online Social Networking: Leakage Prevention and Leakers Detection

## Researchers

Yasmin Bokobza  
yasminbo@post.bgu.ac.il

Prof. Bracha Shapira  
bshapira@bgu.ac.il

Dr. Rami Puzis  
puzis@bgu.ac.il

Prof. Lior Rokach  
liorrk@bgu.ac.il

## Goals

The explosion of online social networking in recent years has damaged organizations due to information leakage by their employees. Employees' social networking activity provides an opportunity for adversaries to extract information from Online Social Networks (OSNs) that may not appear on the official organizational website. This new reality has forced organizations to recognize the need to pay closer attention to the use of OSNs by their employees. For large organizations with thousands of employees, analysis of the content to which all employees are exposed or distribute is unfeasible. Detecting private information leakage and identifying the employees at the source of the leaks are very complex tasks.

## Description

In this research, we detect unintentional private information leakage by employees on social networking sites as soon as possible by intelligently selecting organization member profiles and monitoring their activity. We propose and evaluate efficient SN crawling strategies that are based on topology and central features of the users, such as the number of followers and page rank scoring.

Once we detect private information leakage, we would like to identify the employee at the source of the leak. In order to detect a leaker, the friends of a monitored profile must be inspected. This inspection includes analyzing the content to which these friends were exposed or distributed. We propose and evaluate strategies for identifying the employees who leak private information, with emphasis on high precision and minimal time and effort.

## Results

We used three datasets: Flickr, Digg and Ning. In our evaluation we identified communities within the social networks and referred to each community as an organization. To identify the communities, we used the label propagation algorithm, since it runs in nearly-linear time, allowing for the analysis of large OSN data sets.

Our results show that by monitoring the activity of the users with the highest page rank scoring we can detect more leaks with less effort, by monitoring fewer users. Moreover, by inspecting the friends of a monitored profile with the highest number of followers, we identify more leakers in minimal time and effort.