

Propuesta Proyecto de Grado

Nicolás Acevedo Sandoval
Departamento de Ingeniería de Sistemas y Computación
Universidad de los Andes
Bogotá, Colombia
n.acevedos@uniandes.edu.co

Bibiana Gamba Sabogal
Departamento de Ingeniería de Sistemas y Computación
Universidad de los Andes
Bogotá, Colombia
b.gamba10@uniandes.edu.co

Abstract—A lo largo de este *paper* se busca realizar un análisis de redes para poder entender el comportamiento de los diferentes contribuyentes de repositorios de GitHub. El principal objetivo es entender cómo estas relaciones tienen una influencia en los proyectos de ingeniería de software. De esta forma, se realizará el análisis sobre un repositorio teniendo en cuenta como relación principal de los usuarios la contribución a un mismo archivo. Al final, se utilizan diferentes métricas para poder encontrar *clusters*, conectividad y centralidad de la red.

Palabras clave—*repositorio, contribuyentes, clase, análisis de redes, red social, ingeniería de software, GitHub*

I. PROBLEMA/FENÓMENO A ANALIZAR

Actualmente, los repositorios son una herramienta muy útil en la ejecución de diferentes proyectos de desarrollo de software. Es por esta razón, que la ingeniería de software le ha dado una especial importancia al estudio de los repositorios para brindar mejores soluciones en términos de calidad, costo y agilidad. Una de las herramientas para el manejo de repositorios más utilizadas a nivel mundial es GitHub, donde desarrolladores y otros contribuyentes de proyectos pueden interactuar con el objetivo de entregar mejores soluciones. Teniendo esto en cuenta, las relaciones que se dan a través de la aplicación se vuelven fundamentales para entender dinámicas del proceso y proponer mejoras. A partir de estas interacciones, utilizar análisis de las redes de contribuyentes del proyecto, puede ser un factor diferencial que permita alcanzar el éxito a un proyecto. Además, de encontrar comportamientos de la red que hacen que se logren sus objetivos satisfactoriamente.

II. ESTADO DEL ARTE

A. Estudios empíricos previos

Existen algunos estudios alrededor del análisis de las redes de GitHub. Dentro de los estudios previos acerca del tema, se encuentra el estudio de la red social de desarrollo de software en GitHub. En este *paper* se realiza especialmente un análisis de redes sociales. Se genera una red de los usuarios de algunos repositorios y se busca encontrar propiedades de una red *scale-free*. Esto quiere decir, que buscan esos nodos que interconectan dos o más grupos dentro de la red, lo cual les da una gran importancia. Para ello utilizan métricas de estas redes, en especial buscan bajas puntuaciones en cuanto a densidad, transitividad y centralización. [2]

Por otro lado, se realizó un análisis de los desarrolladores que más influencia tienen en la red de GitHub. Se tomaron 16 millones de usuarios de un periodo de aproximadamente 9 años. A partir de esta información se realizan conexiones entre los

usuarios dependiendo de los seguidores, estrellas, *forks* y actividad en la plataforma. Como resultado, encuentran los diferentes tipos de relaciones y a partir de esta combinación de atributos, la influencia de los desarrolladores dentro de la red. Lo anterior, con diferentes métricas y desde diferentes perspectivas. [1]

Se puede observar, que a pesar de que hay diferentes estudios alrededor de la red social de GitHub existen oportunidades de analizar esta red desde otras perspectivas como lo es la ingeniería de software.

B. Brechas conceptuales

Con respecto a las limitaciones que tiene el proyecto se puede encontrar el hecho de realizar el análisis para una gran cantidad de repositorios. Inicialmente, el estudio se hará para uno o dos repositorios por la complejidad de la formación de la red. Por otro lado, después de observar análisis previos encontramos oportunidades de análisis sobre los repositorios. A diferencia de otros estudios lo que se propone es observar el papel de la ingeniería de software dentro de los repositorios de GitHub. Lo anterior, hacer referencia a las interacciones de los contribuyentes a los mismos archivos o clases para entender el flujo de información dentro de un repositorio.

III. OBJETIVOS

El objetivo general del proyecto es realizar un análisis exploratorio de las relaciones de los contribuyentes de proyectos *open-source* en GitHub.

A. Objetivos específicos

- Investigar sobre el estado del arte en términos de minería de repositorios y análisis de redes sociales para tareas de ingeniería de software.
- Usar técnicas de minería de repositorios, con el objetivo de obtener información relevante para observar las relaciones entre las personas involucradas en un proyecto de software
- Construir una red social que permita identificar relaciones entre los colaboradores de un proyecto que permita mejorar el proceso de desarrollo.
- Proponer mejoras al proceso de desarrollo de software

IV. LISTADO DE ACTIVIDADES

Las siguientes son las actividades que se cumplirán a lo largo del semestre para lograr todos los objetivos del proyecto.

A. Definición de actores importantes y sus relaciones

- Identificar las relaciones más importantes entre desarrolladores
- Identificar atributos relevantes para la construcción del modelo

B. Minería de repositorios

- Descarga de la información del repositorio que se va utilizar
- Extracción de información de actores
- Extracción de relaciones

C. Construcción de la red

- Representación de la información en grafos.
- Visualizar la red con sus actores y conexiones en el proyecto de desarrollo de software.

D. Análisis de la red

- Encontrar grupos de colaboración
- Identificar actores claves
- Proponer mejoras para la asignación de nuevos requerimientos, corrección de errores, entre otras.

V. CRONOGRAMA DE ACTIVIDADES

El siguiente es el cronograma con las actividades a realizar a lo largo del semestre. Las actividades están divididas por semanas, desde la quinta del periodo académico.

Actividad	Semana										
	5	6	7	8	9	10	11	12	13	14	15
Definición de actores importantes y sus relaciones											
Minería de repositorios											
Construcción de la red											
Análisis de la red											
Elaboración del documento final											

Fig. 1. Cronograma de actividades para el proyecto de grado

VI. RESULTADOS ESPERADOS

A partir de la investigación y análisis de la red de los contribuyentes de un repositorio de GitHub, se espera encontrar, que las interacciones entre usuarios especialmente al modificar una misma parte de código pueden ayudar a tareas de ingeniería de software. Además, se podrán encontrar los grupos de desarrolladores que se dedican a requerimientos específicos y la interacción entre las diferentes áreas de un proyecto. También, se espera poder identificar los usuarios que tienen mayor influencia en el resto de la red, para así poder conocer que partes del proyecto podrían controlar y afectar más. Al final del proyecto se espera tener una visualización de la red construida, scripts donde se muestre el trabajo realizado para las tareas de minería de repositorios y la implementación de algoritmos de redes sociales en este contexto. Todo esto con el fin de mejorar el proceso de desarrollo de software.

REFERENCIAS

- [1] Hu, Y., Wang, S., Ren, Y., & Choo, K. (2018). User influence analysis for github developer social networks. *Expert Systems with Applications*, 108, 108-118. doi:10.1016/j.eswa.2018.05.002
- [2] Leibzon, W. (2016) Social network of software development at GitHub. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. San Francisco, CA. pp. 1374-1376.