

User Profiles in CQA: A Focus on the Behavior of Question-Asking Users

Gabriela N. Aranda, Valeria Zoratto, and Adriana Cárdena

Universidad Nacional del Comahue, Neuquén, Argentina
GIISCO, Grupo de Investigación en Ingeniería del Software del Comahue
`gabriela.aranda, valeria.zoratto@fi.uncoma.edu.ar`

Resumen In recent years, there has been growing interest in studying the social networks that emerge from user interactions within Community Question Answering (CQA) platforms, particularly with regard to the roles users assume in collaborative environments. However, literature reviews reveal limited attention to the specific analysis of the behavior of users who formulate questions, in contrast to the extensive body of research focused on those who provide answers. In this context, this work presents a study aimed at characterizing the behavior of question-asking users through a classification model based on participation and quality dimensions of their contributions. The analysis was conducted on two datasets derived from real discussion forums, enabling the identification of differentiated behavioral patterns and the evaluation of the model's stability across different temporal contexts.

The results show that most users concentrate their activity in the early stages of participation, progressively decreasing their level of engagement over time. Furthermore, there is a predominance of profiles characterized by low participation and low quality, as well as limited mobility toward higher-performing profiles. These findings contribute to a better understanding of the user lifecycle in CQA communities and provide relevant insights for the design of strategies aimed at fostering sustained participation and improving the quality of contributions.

Keywords: CQA communities, Discussion forums, Information retrieval, User profiles, Question-asking users

Perfiles de usuarios en CQA: un enfoque en el comportamiento de quienes formulan preguntas

Gabriela N. Aranda, Valeria Zoratto, and Adriana Cárdena

Universidad Nacional del Comahue, Neuquén, Argentina
GIISCO, Grupo de Investigación en Ingeniería del Software del Comahue
`gabriela.aranda, valeria.zoratto@fi.uncoma.edu.ar`

Resumen En los últimos años, se ha incrementado el interés en el estudio de las redes sociales que emergen de la interacción entre usuarios en las Comunidades Virtuales de Preguntas y Respuestas (CQA), particularmente en relación con los roles que estos desempeñan en entornos colaborativos. No obstante, las revisiones bibliográficas evidencian una atención limitada al análisis específico del comportamiento de los usuarios que formulan preguntas, en contraste con el amplio desarrollo de estudios centrados en quienes proveen respuestas. En este contexto, se presenta un estudio orientado a caracterizar el comportamiento de los usuarios que realizan preguntas, a partir de un modelo de clasificación basado en dimensiones de participación y calidad de sus contribuciones. El análisis se llevó a cabo sobre dos conjuntos de datos provenientes de foros de discusión reales, lo que permitió identificar patrones de comportamiento diferenciados y evaluar la estabilidad del modelo en distintos contextos temporales.

Los resultados evidencian que la mayoría de los usuarios concentra su actividad en etapas tempranas de participación, disminuyendo progresivamente su nivel de intervención a lo largo del tiempo. Asimismo, se observa una predominancia de perfiles caracterizados por baja participación y baja calidad, así como una baja movilidad hacia perfiles de mayor desempeño. Estos hallazgos contribuyen a una mejor comprensión del ciclo de vida de los usuarios en comunidades CQA y aportan evidencia relevante para el diseño de estrategias orientadas a fomentar la participación sostenida y mejorar la calidad de las contribuciones.

Keywords: Comunidades CQA, Foros de discusión, Recuperación de información, Perfiles de usuario, Usuarios que formulan preguntas

1. Introducción

Las Comunidades Virtuales de Preguntas y Respuestas, conocidas como CQA (del inglés Community Question Answering), constituyen una modalidad ampliamente extendida de búsqueda de información en línea. En estos entornos, los usuarios intentan resolver problemáticas de la vida cotidiana mediante el intercambio colaborativo de conocimiento. Estas plataformas promueven la interacción entre participantes a través de estructuras que facilitan tanto la formulación de preguntas como la provisión de respuestas por parte de la comunidad.

Entre las CQA más difundidas, los foros de discusión se destacan por concentrar información de alto valor para los usuarios, especialmente en la resolución de problemas técnicos recurrentes. En este contexto, en los últimos años ha crecido el interés por el desarrollo de técnicas de recuperación que permitan analizar la información contenida en estos hilos de discusión y potenciar su reutilización mediante herramientas de búsqueda automáticas o semiautomáticas.

En particular, los foros de discusión técnicos constituyen un tipo específico de CQA que se ha consolidado como un repositorio dinámico y actualizado de conocimiento especializado en diversos dominios. La participación activa de los usuarios, basada en el intercambio de experiencias y soluciones ante problemas recurrentes, genera recursos de gran utilidad para la comunidad. Por esta razón, estos entornos han adquirido relevancia en múltiples líneas de investigación orientadas a la aplicación de técnicas de Recuperación de Información (RI) para el análisis y aprovechamiento de sus contenidos.

A partir de una amplia revisión de trabajos relacionados con la participación de usuarios en foros de discusión, se observa una limitada cantidad de estudios centrados específicamente en el análisis del comportamiento de los usuarios que formulan preguntas en foros técnicos. En este contexto, se propone la definición de un modelo de clasificación de perfiles de este tipo de usuarios, el cual es evaluado sobre conjuntos de datos reales provenientes del foro técnico StackOverflow¹, perteneciente a la plataforma StackExchange².

La elección de este foro se fundamenta en su amplio uso dentro de la comunidad de programadores, así como en su pertenencia a la red StackExchange, lo que posibilita la replicabilidad del estudio, con mínimas adaptaciones, en otros foros de la misma plataforma orientados a diversas temáticas, tales como aprendizaje de idiomas, animé, matemática, entre otras.

En este contexto, las principales contribuciones de este trabajo son las siguientes: (i) la definición de un modelo de clasificación orientado específicamente a usuarios que formulan preguntas en comunidades CQA, basado en dimensiones de participación y calidad; (ii) el análisis longitudinal de la evolución de dichos perfiles, considerando tanto su distribución como las transiciones entre ellos a lo largo del tiempo; y (iii) la validación empírica del modelo propuesto sobre datos reales de StackOverflow, utilizando distintas cohortes de usuarios, lo que permite evidenciar patrones consistentes de comportamiento y evaluar la estabilidad del enfoque en diferentes contextos temporales.

El resto del artículo se organiza de la siguiente manera. En la Sección 2 se presenta un análisis de los trabajos relacionados. A continuación, en la Sección 3, se describen las bases de la investigación desarrollada en este estudio. Posteriormente, los resultados obtenidos se exponen en la Sección 4. Finalmente, las conclusiones y las líneas de trabajo futuro se discuten en la sección final.

¹ <https://stackoverflow.com>

² <https://stackexchange.com/>

2. Trabajos relacionados

A partir de una amplia revisión de la literatura sobre usuarios en foros de discusión, se identificaron siete trabajos con puntos de contacto con el objeto de estudio de esta investigación, aunque ninguno lo aborda de manera integral. Con el objetivo de sintetizar sus principales características, dichos trabajos fueron analizados considerando los siguientes ejes comparativos: comportamiento de los usuarios (Comp), relación entre usuarios (Rel), análisis temporal (Temp), tiempo de permanencia o actividad en la comunidad (Perm), estudio de un tipo de usuario particular (UsP), calidad de la pregunta (Cal) y clasificación de usuarios (Clas). Un resumen de este análisis se presenta en la Tabla 1.

Tabla 1. Comparación de trabajos relacionados según enfoque

Autores	Foro	Comp	Rel	Temp	Perm	UsP	Cal	Clas
Pal et al., 2012	StackOverflow	•	•	•		•		
Wang et al., 2013	StackOverflow	•	•					
Furtado et al., 2013	StackExchange	•		•			•	•
Squire y Funkhouser, 2014	StackOverflow						•	
Shoji et al., 2015	Yahoo! Chiebukuro	•		•	•			
Slag et al., 2015	StackOverflow	•			•	•		
Baltadzhieva y Chrupala, 2015	General						•	

Como se observa en la columna *Foro*, la mayoría de los trabajos (cinco de siete) utiliza hilos de discusión de la plataforma StackExchange, probablemente debido a su disponibilidad pública y a sus actualizaciones periódicas. En particular, cuatro de ellos se enfocan en el foro StackOverflow (SO), ampliamente utilizado por la comunidad de programadores y caracterizado por su gran volumen de preguntas y respuestas técnicas. En contraste, solo el trabajo de Baltadzhieva y Chrupala, 2015 analiza características comunes a distintos foros sin centrarse en uno específico.

En cuanto a los enfoques considerados, la mayoría de los estudios (cinco de siete) analiza el comportamiento de los usuarios (Comp), mientras que solo dos abordan la relación entre pares de usuarios (Rel). Por otro lado, el análisis temporal de la participación (Temp) aparece en menos de la mitad de los trabajos (tres de siete). Respecto del estudio de tipos particulares de usuarios (UsP), se destacan los trabajos de Pal et al., 2012, centrado en la evolución de usuarios expertos, y de Slag et al., 2015, orientado al análisis de los denominados *one-day-flies*. A partir de este relevamiento, se identifica una vacancia en estudios enfocados específicamente en el comportamiento y la clasificación de usuarios que formulan preguntas, lo cual motiva la presente investigación.

Finalmente, en relación con la clasificación de usuarios (Clas), únicamente el trabajo de Furtado et al., 2013 propone una tipología compuesta por diez perfiles de usuario. Sin embargo, solo uno de ellos corresponde específicamente a usuarios que formulan preguntas —denominado “Preguntador Experto” (EQ)—, lo que refuerza la pertinencia del enfoque adoptado en este trabajo.

3. Propuesta de clasificación de usuarios que preguntan

Para la definición de una propuesta de clasificación de los usuarios que formulan preguntas en foros de discusión técnicos, se toma como referencia el estudio de Furtado et al., 2013, en el cual se identifican diez perfiles de usuario clasificados a partir de métricas calculadas sobre su historial completo de actividad en la plataforma.

Dado que el presente trabajo se centra exclusivamente en los usuarios que preguntan, las métricas definidas en Furtado et al., 2013 fueron adaptadas de la siguiente manera:

- *Métrica de motivación*: se define a partir del número de preguntas publicadas y de la duración de la actividad del usuario, medida como la cantidad de días activos desde la fecha de creación de la cuenta hasta la última participación registrada (ya sea pregunta o respuesta).
- *Métrica de habilidad*: se calcula considerando el número de veces que las preguntas han sido marcadas como favoritas (indicador de utilidad percibida) y el balance de votos recibidos (diferencia entre votos positivos y negativos). Este valor, estimado para cada usuario, representa la utilidad promedio de sus preguntas.

A partir de estas métricas, se establecen cuatro perfiles de usuarios que preguntan, considerando las siguientes dimensiones:

- *Participación*: cantidad total de preguntas realizadas por el usuario.
- *Calidad de las preguntas*: medida que refleja la percepción de la comunidad respecto de la utilidad de las preguntas de un usuario. Para su cálculo, se consideran los votos recibidos (positivos y negativos) y la cantidad de veces que las preguntas han sido marcadas como favoritas.

A partir de estos valores, agregados y promediados sobre el total de preguntas realizadas por el usuario, se obtiene una medida de calidad promedio por usuario, definida mediante la ecuación 1:

$$promU = \frac{\sum p}{totalP} \quad (1)$$

$$p = (v_p - v_n) + m_f \quad (2)$$

donde:

- $promU$: calidad promedio del usuario
- p : calidad de cada pregunta (ecuación 2)
- $totalP$: total de preguntas realizadas
- v_p : votos positivos
- v_n : votos negativos
- m_f : cantidad de veces que la pregunta fue marcada como favorita

Para cada una de las medidas se definieron dos niveles (alto/bajo), tomando como referencia el valor medio de dicha medida calculado sobre el conjunto de usuarios bajo estudio que realizaron al menos una pregunta durante el período analizado.

En particular, en el caso de la *participación*, esta se considera *alta* cuando la cantidad de preguntas realizadas por un usuario es superior a la media del conjunto total de usuarios, y *baja* en caso contrario. De manera análoga, para la calidad percibida por la comunidad respecto de las preguntas de un usuario, se toma como referencia el valor medio de la utilidad de sus preguntas en relación con el promedio global del conjunto de usuarios.

A partir de estas definiciones, se establecen cuatro perfiles de usuarios que formulan preguntas en foros de discusión, tal como se muestra en la Tabla 2. En este esquema, el Perfil 1 corresponde a usuarios con alta participación y alta calidad (valores superiores a la media en ambas dimensiones), el Perfil 2 a usuarios con alta calidad y baja participación, el Perfil 3 a usuarios con baja calidad y alta participación, y el Perfil 4 a usuarios con baja participación y baja calidad.

Tabla 2. Perfiles según participación del usuario y calidad de las preguntas

	Alta Calidad	Baja Calidad
Alta Participación	Perfil 1	Perfil 2
Baja Participación	Perfil 3	Perfil 4

En las secciones siguientes se presenta la aplicación de esta clasificación sobre usuarios correspondientes a dos casos de estudio.

4. Resultados obtenidos

Con el objetivo de analizar la aplicabilidad de la propuesta de clasificación de usuarios que formulan preguntas, se desarrolló una herramienta destinada a la recuperación y procesamiento de información de distintos conjuntos de usuarios de StackOverflow, así como de los posts generados por estos.

Dicha herramienta permitió realizar tareas de filtrado inicial de datos, agrupamiento y cálculo de diversas métricas relevantes, tales como: cantidad total de posts generados por cada usuario, distribución de dichos posts en intervalos temporales de 12 meses desde la fecha de registración, cantidad de veces que un post fue marcado como favorito por otros usuarios, y balance entre votos positivos y negativos recibidos por cada publicación, entre otras. Estas métricas constituyen la base para el cálculo de las dimensiones de participación y calidad definidas en la sección anterior.

La herramienta fue aplicada sobre datos reales obtenidos de la base de datos de StackOverflow, correspondientes a la versión del 03/06/2019. A partir de este conjunto de datos, se definieron dos casos de estudio con el fin de evaluar

la robustez de la propuesta en contextos temporales y volúmenes de usuarios diferentes.

El primer caso de estudio (caso de estudio 1) corresponde a un análisis preliminar sobre un conjunto de 250 usuarios registrados durante el mes de creación del foro StackOverflow (agosto de 2008). Este grupo permite observar el comportamiento de usuarios pioneros dentro de la plataforma.

El segundo caso de estudio (caso de estudio 2) está conformado por 11.298 usuarios registrados a lo largo del año 2011, lo que permite analizar patrones de comportamiento en una etapa de mayor madurez de la comunidad.

En ambos casos, se estableció como criterio de inclusión que los usuarios hubieran registrado su último acceso al foro durante el año 2018 (atributo *lastAccessDate* en la tabla *Users*). Este criterio permite comparar el comportamiento de usuarios que han permanecido activos durante un período de tiempo similar. En este trabajo, se considera como usuarios activos a aquellos que han mantenido algún tipo de interacción con la plataforma, ya sea mediante la formulación de preguntas, la publicación de respuestas, la realización de comentarios o la consulta de contenidos.

Adicionalmente, se analizó la evolución temporal de las métricas definidas, con el objetivo de identificar posibles cambios en los patrones de participación y calidad a lo largo del tiempo. Este enfoque permite no solo caracterizar los perfiles de usuario en términos estáticos, sino también comprender su dinámica de evolución dentro de la comunidad.

A continuación, se presentan los resultados obtenidos para ambos casos de estudio.

4.1. Distribución de usuarios por perfil

En la Figura 1 se presenta la distribución de los usuarios de ambos casos de estudio en los cuatro perfiles propuestos. A la izquierda se muestran los resultados correspondientes al caso de estudio 1 (250 usuarios), mientras que a la derecha se observan los del caso de estudio 2.

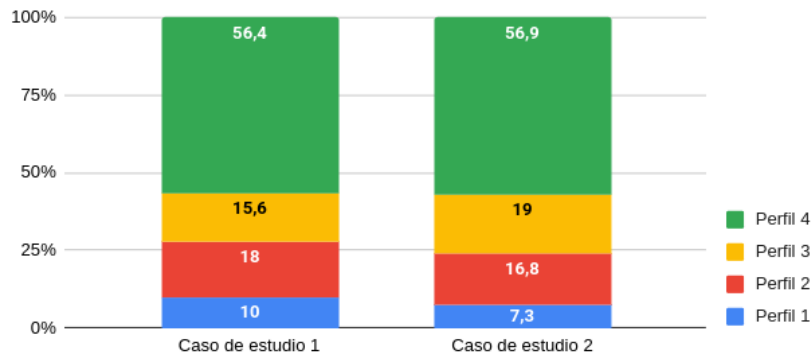


Figura 1. Distribución de usuarios según perfil

En el caso de estudio 1, se observa que la mayoría de los usuarios (56,4 %) se concentra en el Perfil 4, caracterizado por una baja participación (por debajo de la media de 14 preguntas) y una baja calidad promedio de las preguntas (inferior al valor medio de 32,97). Del 43,6 % restante, un 28 % de los usuarios presenta alta participación (Perfiles 1 y 2), aunque solo un 10 % del total combina alta participación con alta calidad (Perfil 1). Por otro lado, el 15,6 % de los usuarios se ubica en el Perfil 3, correspondiente a aquellos con baja participación pero alta calidad en sus preguntas.

En el caso de estudio 2, la media de participación es de 10 preguntas por usuario y la calidad promedio es de 3. Al igual que en el caso anterior, la mayor proporción de usuarios (56,9 %) se clasifica en el Perfil 4. Del 43,1 % restante, únicamente un 7,3 % corresponde al Perfil 1 (alta participación y alta calidad), mientras que el 35,8 % se distribuye entre los Perfiles 2 y 3, es decir, usuarios con alta participación pero baja calidad, o bien con baja participación y alta calidad.

En términos generales, los resultados evidencian un patrón consistente en ambos casos de estudio: la mayoría de los usuarios presenta bajos niveles tanto de participación como de calidad, mientras que el grupo de usuarios con desempeño destacado (Perfil 1) constituye una minoría. Asimismo, la similitud en las distribuciones sugiere la estabilidad del modelo propuesto frente a conjuntos de datos de distinto tamaño y contexto temporal.

4.2. Pertenencia de usuarios por perfil a lo largo de los años

Continuando con el análisis, se estudió la pertenencia de los usuarios de los distintos perfiles y su evolución a lo largo del tiempo dentro de la comunidad. Para ello, en la Tabla 3 se presenta el porcentaje de usuarios correspondientes a cada perfil en períodos consecutivos de 12 meses para el caso de estudio 1 (2008–2018), mientras que en la Tabla 4 se muestran los resultados para el caso de estudio 2 (2011–2018).

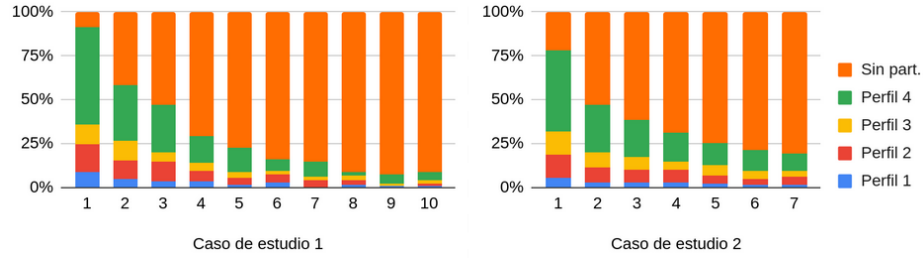
Tabla 3. Porcentaje de usuarios por periodo según perfil (Caso de estudio 1)

Periodos	1	2	3	4	5	6	7	8	9	10
Perfil 1	8,8 %	4,4 %	3,2 %	3,2 %	1,6 %	2,4 %	0 %	1,2 %	0,8 %	0,4 %
Perfil 2	16 %	10,8 %	11,6 %	6 %	4 %	4,8 %	4 %	2,8 %	0 %	1,6 %
Perfil 3	11,2 %	11,6 %	4,8 %	5 %	3 %	2 %	2 %	2,4 %	1,2 %	2 %
Perfil 4	55,6 %	31,6 %	27,2 %	15 %	14 %	6,8 %	8,8 %	2 %	5,6 %	4,8 %
Sin participación	8,4 %	41,6 %	53,2 %	70,8 %	78 %	84 %	85,2 %	91,6 %	92,4 %	91,2 %

En relación con la permanencia de la actividad según perfil, en el caso de estudio 1 se observa que, en la mayoría de los períodos, los usuarios activos se concentran principalmente en el Perfil 4 (baja participación y baja calidad de preguntas). Una excepción se presenta en el período 8, donde la distribución de usuarios resulta más equilibrada entre los distintos perfiles, con una leve predominancia del Perfil 2 (alta participación y baja calidad).

Tabla 4. Porcentaje de usuarios por periodo según perfil (Caso de estudio 2)

Periodo	1	2	3	4	5	6	7
Perfil 1	5,5 %	2,8 %	2,7 %	2,4 %	2,1 %	1,6 %	1,6 %
Perfil 2	13,4 %	8,7 %	7,0 %	7,3 %	4,7 %	2,8 %	4,4 %
Perfil 3	13,1 %	8,1 %	7,3 %	5,0 %	6,1 %	5,1 %	3,2 %
Perfil 4	45,9 %	27,2 %	21,7 %	16,7 %	12,4 %	11,5 %	9,9 %
Sin participación	22,1 %	53,3 %	61,2 %	68,6 %	74,8 %	79,0 %	80,9 %

**Figura 2.** Distribución de usuarios según perfil por periodo

Por el contrario, el Perfil 1 (alta participación y alta calidad) es consistentemente el menos representado a lo largo de todos los períodos. En particular, en el período 7 no se registran usuarios en este perfil, y en los períodos finales su proporción se mantiene por debajo del 1 %. Asimismo, se evidencia un marcado incremento en la proporción de usuarios sin participación a partir del segundo período. Este grupo supera el 50 % desde el tercer período y alcanza valores superiores al 90 % a partir del octavo período, lo que indica una fuerte disminución de la actividad a lo largo del tiempo.

En el caso de estudio 2, se identifican patrones similares. En el primer período, cerca de la mitad de los usuarios (45,9 %) se ubica en el Perfil 4, mientras que el 22,1 % no presenta participación. Los perfiles 2 y 3 muestran proporciones similares (alrededor del 13 %), y el Perfil 1 vuelve a ser el menos representado (5,5 %).

A partir del segundo período, se observa un aumento significativo de usuarios sin participación (de 22,1 % a 53,3 %), tendencia que se mantiene en los períodos siguientes. Entre los usuarios activos, el Perfil 4 continúa siendo predominante, mientras que los perfiles 2 y 3 presentan valores similares entre sí, y el Perfil 1 se mantiene como el menos frecuente.

En términos generales, tal como se muestra en la Figura 2, los resultados muestran una tendencia consistente en ambos casos de estudio: a medida que transcurre el tiempo, se incrementa considerablemente la proporción de usuarios sin actividad, mientras que entre los usuarios activos predominan aquellos con baja participación y baja calidad (Perfil 4). Asimismo, los perfiles intermedios (2 y 3) presentan comportamientos similares, y el grupo de usuarios con alta

participación y alta calidad (Perfil 1) constituye una minoría persistente en todos los períodos analizados.

4.3. Cambio entre perfiles

Dado que el análisis previo presenta una visión global del comportamiento de los usuarios a lo largo del tiempo, resulta de interés analizar la dinámica de cambio entre perfiles en períodos consecutivos.

Para ello, se definieron ventanas temporales de 12 meses, medidas desde la fecha de registración del usuario. En cada ventana, se calcularon las métricas de participación y calidad considerando únicamente la actividad dentro de dicho intervalo, y se asignó a cada usuario el perfil correspondiente en cada período.

Con el objetivo de analizar las transiciones entre períodos consecutivos, se incorporó además un estado *inactivo*, que agrupa a los usuarios sin participación durante una ventana completa de 12 meses. Este estado no constituye un perfil propiamente dicho, ya que la ausencia de actividad impide evaluar las dimensiones de participación y calidad.

A partir de esta definición, se calcularon las probabilidades de transición entre perfiles (y hacia el estado inactivo), considerando la frecuencia de cambios entre cada par de estados. Los resultados para el caso de estudio 1 se presentan en la Tabla 5.

Tabla 5. Probabilidad de cambio de perfil en periodos consecutivos (Caso 1)

	Perfil 1	Perfil 2	Perfil 3	Perfil 4	Inactivo
Perfil 1	10,9 %	21,9 %	17,2 %	20,3 %	29,7 %
Perfil 2	0 %	30,6 %	7,5 %	30,6 %	31,3 %
Perfil 3	1,8 %	6,5 %	8,3 %	24,1 %	59,3 %
Perfil 4	1,7 %	7,2 %	6,3 %	24,3 %	60,5 %
Inactivo	0,7 %	1,5 %	1,9 %	7,1 %	88,8 %

En términos generales, se observa que, independientemente del perfil inicial, la transición más probable en el período siguiente es hacia el estado inactivo, lo que evidencia una alta tasa de abandono de la participación en la comunidad. Asimismo, una vez que un usuario se encuentra en estado inactivo, existe una alta probabilidad de que permanezca en dicho estado (88,8 %), siendo muy poco frecuente su retorno a perfiles activos.

El análisis por perfil permite identificar los siguientes comportamientos. En el caso del Perfil 1 (alta participación y alta calidad), los usuarios presentan una mayor diversidad de transiciones, siendo más probable que disminuyan la calidad (pasando al Perfil 2) o ambas dimensiones (Perfil 4) que permanecer en el mismo perfil. No obstante, este perfil es el que presenta menor probabilidad de transición hacia la inactividad.

Para los usuarios del Perfil 2 (alta participación y baja calidad), se observa una baja probabilidad de mejorar la calidad de sus aportes, mientras que es

igualmente probable que se mantengan en el mismo perfil o que disminuyan su participación, pasando al Perfil 4.

En los perfiles 3 y 4, las probabilidades de transición son similares, predominando los cambios hacia el Perfil 4 o hacia el estado inactivo, lo que sugiere una tendencia a la disminución progresiva de la actividad.

Desde una perspectiva inversa (por columnas), se observa que el Perfil 1 presenta una baja capacidad de atracción desde otros perfiles, mientras que el Perfil 4 actúa como un estado de convergencia hacia el cual tienden a transitar usuarios provenientes de distintos perfiles. Asimismo, se identifica que el Perfil 4 es el destino más probable para los usuarios que retoman la actividad luego de un período de inactividad.

De manera análoga, se realizó el análisis para el caso de estudio 2, cuyos resultados se presentan en la Tabla 6.

Tabla 6. Probabilidad de cambio de perfil en periodos consecutivos (Caso 2)

	Perfil 1	Perfil 2	Perfil 3	Perfil 4	Inactivo
Perfil 1	18 %	22,8 %	12,3 %	21,8 %	25,1 %
Perfil 2	8,1 %	30 %	7,11 %	28,2 %	26,6 %
Perfil 3	3,4 %	5 %	9,4 %	20,4 %	61,8 %
Perfil 4	2 %	6 %	6,7 %	23,4 %	61,9 %
Inactivo	0,7 %	2,1 %	4,5 %	11,8 %	80,9 %

Los resultados obtenidos muestran patrones consistentes con el caso de estudio anterior. En particular, se destaca que:

- Los usuarios del Perfil 1 presentan una alta probabilidad de degradar su desempeño, ya sea en términos de calidad (Perfil 2) o en ambas dimensiones (Perfil 4), siendo menos frecuente su permanencia en el mismo perfil.
- En el Perfil 2, la permanencia es el comportamiento más probable, seguido por la transición hacia el Perfil 4, mientras que la mejora en calidad resulta poco frecuente.
- Los usuarios del Perfil 3 presentan baja estabilidad y una alta tendencia a disminuir su participación o a volverse inactivos.
- El Perfil 4 se caracteriza por una alta persistencia y una muy baja probabilidad de mejora hacia otros perfiles, siendo además uno de los principales estados previos a la inactividad.
- Finalmente, los usuarios inactivos presentan una muy baja probabilidad de reactivarse, manteniéndose en dicho estado en más del 80 % de los casos.

En la Figura 3 se presenta el diagrama de mapa de calor de ambos casos de estudio, donde los colores permiten observar claramente la similitud entre los valores de probabilidad obtenidos para el cambio de perfil entre un periodo de tiempo t y su periodo consecutivo $(t+1)$, en todas las intersecciones. En conjunto, los resultados obtenidos evidencian una dinámica en la que la mayoría

de los usuarios tiende a reducir progresivamente su nivel de actividad, con escasas transiciones hacia perfiles de mayor calidad o participación, lo que refuerza la importancia de identificar tempranamente patrones de comportamiento que permitan fomentar una participación sostenida en la comunidad.

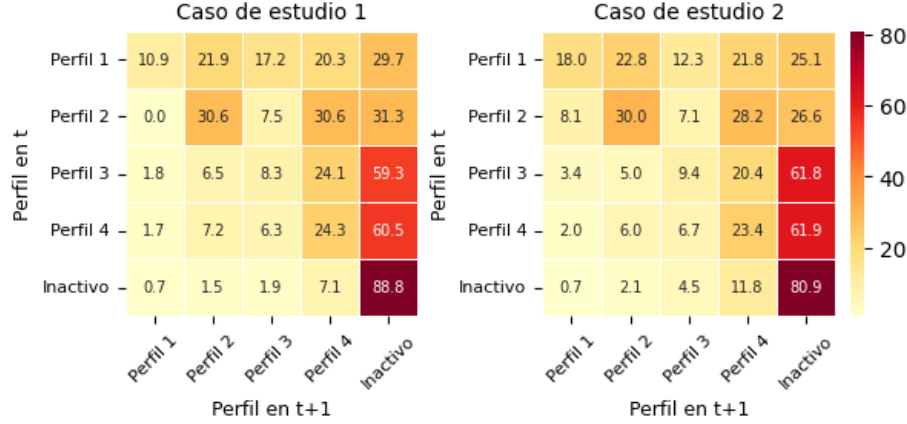


Figura 3. Probabilidad de cambio de perfil en periodos consecutivos

5. Discusión sobre los resultados obtenidos

A modo de resumen de los análisis presentados en las secciones anteriores para ambos casos de estudio, podemos decir que en forma general se ha observado que:

- El grupo mayoritario de usuarios en ambos casos de estudio fue el Perfil 4 (usuarios cuyas preguntas fueron percibidas como de menor calidad y con poca participación), siendo el Perfil 2 (baja calidad, alta participación) el segundo grupo mayoritario en el caso de estudio 1 y el Perfil 3 (alta calidad, baja participación) en el caso de estudio 2, aunque los porcentajes entre los dos perfiles son muy similares en ambos casos de estudio.
- El grupo de usuarios de alta calidad y alta participación (Perfil 1) es el grupo minoritario en todos los periodos en ambos casos de estudio, siendo igualmente probable que en la ventana de tiempo siguiente los usuarios pasen al Perfil 2 o Perfil 4, es decir, que disminuyan la calidad o disminuyan la participación también.
- Para los usuarios del Perfil 2 es igualmente probable que se mantengan en el mismo perfil o que disminuyan la cantidad de preguntas que realizan, pasando a Perfil 4.
- Para los usuarios del Perfil 3, lo más probable es que mantengan la baja participación y bajen también su calidad, pasando a Perfil 4, a que perma-

nezcan en el mismo perfil, siendo la probabilidad de que esto ocurra menor al 10 %.

- Los usuarios del Perfil 4 no suelen mejorar ni su participación ni la calidad de sus preguntas, y suelen mantenerse en el mismo perfil o pasar a ser Inactivos.
- Finalmente, el mayor porcentaje de usuarios inactivos que vuelven a la participación, pasan a formar parte del Perfil 4.

6. Conclusiones y trabajo futuro

En este trabajo se analizaron distintos aspectos relacionados con la evolución de la actividad de los usuarios que formulan preguntas en foros de discusión técnicos, tomando como caso de estudio la comunidad de StackOverflow y considerando dos cohortes de usuarios correspondientes a distintos períodos de incorporación al sitio. Los resultados obtenidos evidencian que la mayor parte de los usuarios concentra su actividad en las etapas iniciales de participación, disminuyendo progresivamente su nivel de intervención a lo largo del tiempo.

Asimismo, la aplicación del modelo de clasificación propuesto permitió identificar diferentes perfiles de usuarios que preguntan, observándose un claro predominio de aquellos caracterizados por baja participación y baja calidad percibida. En este sentido, el análisis de las transiciones entre perfiles pone de manifiesto una dinámica de comportamiento en la que los usuarios tienden a mantener o reducir su nivel de desempeño, siendo poco frecuentes las mejoras sostenidas en términos de participación y calidad.

En conjunto, estos hallazgos sugieren la existencia de un patrón de participación decreciente y baja movilidad hacia perfiles de mayor calidad, lo que constituye un aspecto relevante para la comprensión del ciclo de vida de los usuarios en comunidades CQA. En particular, estos resultados pueden resultar de utilidad para el diseño de estrategias orientadas a fomentar la participación sostenida, mejorar la calidad de las contribuciones y detectar tempranamente posibles situaciones de abandono. Asimismo, estos resultados pueden ser de utilidad para el diseño de mecanismos de recomendación, estrategias de retención de usuarios y sistemas de apoyo a la moderación en comunidades CQA.

Como limitación, el modelo propuesto se basa en métricas agregadas que no consideran el contenido semántico de las preguntas, lo cual podría influir en la caracterización de los perfiles. Asimismo, el análisis se restringe a una única plataforma, lo que podría condicionar la generalización de los resultados.

Como líneas de trabajo futuro, se propone ampliar el análisis incorporando nuevas cohortes de usuarios correspondientes a distintos períodos de inscripción, con el objetivo de evaluar la estabilidad de los patrones observados en diferentes contextos temporales. Asimismo, resulta de interés considerar factores adicionales, tales como la temática de las preguntas, la experiencia previa de los usuarios o su nivel de participación en otras actividades de la plataforma (por ejemplo, respuestas, comentarios o ediciones), que podrían aportar evidencia complementaria para una caracterización más integral de los perfiles identificados.

Agradecimiento

Este trabajo se enmarcó en el subproyecto *Reuso de conocimiento en Foros de Discusión Técnicos - Parte II*, del Programa 04/F009 “Desarrollo de Software Basado en Reuso - Parte II” (2017-2021) y en el Proyecto 04/F018 *Reuso de Información en Comunidades Virtuales* (2022-2026), desarrollados por el Grupo GIISCo de la Facultad de Informática de la Universidad Nacional del Comahue, Neuquén, Argentina.

Referencias

- Baltadzhieva, A., & Chrupala, G. (2015). Question Quality in Community Question Answering Forums: a survey. *SIGKDD Explor.*, 17(1), 8-13. <https://doi.org/10.1145/2830544.2830547>
- Furtado, A., Andrade, N., Oliveira, N., & Brasileiro, F. V. (2013). Contributor profiles, their dynamics, and their importance in five q&a sites. En A. S. Bruckman, S. Counts, C. Lampe & L. G. Terveen (Eds.), *Computer Supported Cooperative Work, CSCW 2013, San Antonio, TX, USA, February 23-27, 2013* (pp. 1237-1252). ACM. <https://doi.org/10.1145/2441776.2441916>
- Pal, A., Chang, S., & Konstan, J. A. (2012). Evolution of Experts in Question Answering Communities. En J. G. Breslin, N. B. Ellison, J. G. Shanahan & Z. Tufekci (Eds.), *Proceedings of the Sixth International Conference on Weblogs and Social Media, Dublin, Ireland, June 4-7, 2012*. The AAAI Press. <https://doi.org/10.1609/icwsm.v6i1.14262>
- Shoji, Y., Fujita, S., Tajima, A., & Tanaka, K. (2015). Who Stays Longer in Community QA Media? - User Behavior Analysis in cQA -. En T. Liu, C. N. Scollon & W. Zhu (Eds.), *Social Informatics - 7th International Conference, SocInfo 2015, Beijing, China, December 9-12, 2015, Proceedings* (pp. 33-48, Vol. 9471). Springer. https://doi.org/10.1007/978-3-319-27433-1_3
- Slag, R., de Waard, M., & Bacchelli, A. (2015). One-Day Flies on StackOverflow - Why the Vast Majority of StackOverflow Users Only Posts Once. En M. D. Penta, M. Pinzger & R. Robbes (Eds.), *12th IEEE/ACM Working Conference on Mining Software Repositories, MSR 2015, Florence, Italy, May 16-17, 2015* (pp. 458-461). IEEE Computer Society. <https://doi.org/10.1109/MSR.2015.63>
- Squire, M., & Funkhouser, C. (2014). A Bit of Code: How the Stack Overflow Community Creates Quality Postings. *47th Hawaii International Conference on System Sciences, HICSS 2014, Waikoloa, HI, USA, January 6-9, 2014*, 1425-1434. <https://doi.org/10.1109/HICSS.2014.185>
- Wang, S., Lo, D., & Jiang, L. (2013). An empirical study on developer interactions in StackOverflow. En S. Y. Shin & J. C. Maldonado (Eds.), *Proceedings of the 28th Annual ACM Symposium on Applied Computing, SAC '13, Coimbra, Portugal, March 18-22, 2013* (pp. 1019-1024). ACM. <https://doi.org/10.1145/2480362.2480557>