Sự phát triển nhanh chóng của mạng xã hội như Twitter1 hay Facebook2 thu hút một số lượng lớn người dùng tham gia kết nối, chia sẻ và giao tiếp với nhau. Đặc biệt người gửi thư rác ngày càng nhằm mục tiêu tới hệ thống Web xã hội, họ tiến hành các cuộc tấn công lừa đảo, phần mềm độc hại, và tin nhắn rác “spam” thương mại. Theo báo cáo thống kê của Nexgate3, trong nửa đầu năm 2013 có một sự tăng trưởng 355% thư rác trên mạng xã hội và Facebook có số lượng các cuộc tấn công lừa đảo gấp bốn lần so với các mạng truyền thông xã hội khác. Do đó một vấn đề đặt ra là nhận diện các tài khoản spam và đưa ra giải pháp ngăn chặn spam. Chủ đề nghiên cứu nhận dạng tài khoản spam trên mạng xã hội thu hút sự quan tâm của cộng đồng nghiên cứu trong nửa thập kỉ trở lại đây. Có nhiều công trình nghiên cứu về các giải pháp phát hiện spam được công bố, chẳng hạn như Benevenuto và cộng sự, Kyumin Lee và cộng sự, Hongyu Gao và cộng sự, 2010