models, such as Susceptible-Infected-Susceptible (SIS), Susceptible-Infected-Recovered (SIR) [20, 21] and random walk [22–24]. At the same time, few works have been done to reveal the coupled dynamics of both the structure and the diffusion of online social networks [25, 26]. To meet this critical challenge, in this paper, we aim to investigate the role of weak ties in the information diffusion in online social networks.

By monitoring the dynamics of

$$\bar{S} = \sum_{S<S_{\max}} \frac{nS^2}{N}, \tag{2}$$

where n is the number of connected clusters with S nodes, and N is the size of the network, a phase transition was found in the mobile communication network during the removal of weak ties first [19]. We find that this phase transition is pervasive in online social networks, which implies that weak ties play a special role in the structure of the network. This interesting finding inspires us to investigate the role of weak ties in the information diffusion. To this end, we propose a model $ID(\alpha, \beta)$ to characterize the mechanism of the information diffusion in online social networks and associate the strength of ties with the process of spread. Through the simulations on large-scale real-world data sets, we find that selecting weak ties preferentially to republish cannot make the information diffuse quickly, while the random selection can. Nevertheless, further analysis and experiments show that the coverage of the information will drop substantially during the removal of weak ties even for the random diffusion case. So we conclude that weak ties play a subtle role in the information diffusion in online social networks. We also discuss their potential use for the information diffusion control practices.

The rest of this paper is organized as follows. Section II introduces the data sets used in this paper. In Section III, we study the structural role of weak ties. The model $ID(\alpha, \beta)$ is proposed in Section IV, and the role of weak ties in the information diffusion is then investigated. Section V discusses the possible uses of weak ties in the control of the virus spread and the private information diffusion. Finally, we give a brief summary in Section VI.

II. DATA SETS

We use two data sets in this paper, i.e., `YouTube` and `Facebook` in New Orleans. `YouTube` is a famous video sharing site, and `Facebook` is the most popular online social site which allows users to create friendships with other users, publish blogs, upload photos, send messages, and update their current states on their profile pages. All these sites have some privacy control schemes which control the access to the shared contents. The data set of `YouTube` includes user-to-user links crawled from `YouTube` in 2007 [8]. The data set of `Facebook` contains a list of all the user-to-user links crawled from the New Orleans regional network in `Facebook` during December 29th, 2008 and January 3rd, 2009 [14]. In both two data sets, we treat the links as undirected.

TABLE I: Data Sets

Data set	$\lvert V\rvert$	$\lvert E\rvert$
`YouTube`	1134890	2987624
`Facebook`	63392	816886

In these data sets, each node represents a user, while a tie between two nodes means there is a friendship between two users. In general, creating a friendship between two users always needs mutual permission. So we can formalize each data set as an undirected graph $G(V, E)$, where V is the set of nodes and E is the set of ties. We use $|V|$ to denote the size of the network, and $|E|$ to denote the size of ties. Some characteristics of the data sets are shown in Table I. The *Cumulative Distribution Function*(CDF) of the strength of ties is shown in Fig. 1.

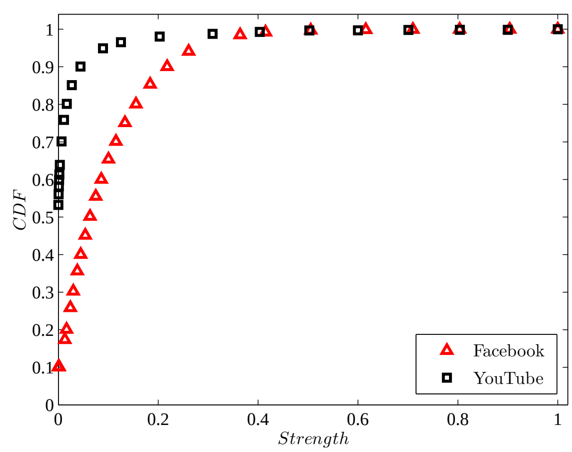

FIG. 1: (Color online) CDF of the strength of ties.

As we know, online social networks are divided into two types: knowledge-sharing oriented and networking oriented [15]. For the data sets we use, `YouTube` belongs to the former, while `Facebook` belongs to the latter, both of which are scale-free networks.

III. STRUCTURAL ROLE OF WEAK TIES

In this section, we study the structural role of weak ties. As shown in Fig. 2a and Fig. 2c, we find a phase transition (characterized by $\bar{S}$) similar to the one in [19] in online social networks during the removal of weak ties first. This phase transition, however, disappears if we remove the strong ties first. Furthermore, it is also found in Fig. 2b and Fig. 2d that the relative size of giant con-