

SUMMARY

[0007] There is provided in accordance with an aspect of the present disclosure a system for identifying attacks against a customer in a social network. The system includes a conversation monitor, a user type detector and an attack identifier. The conversation monitor
5 discovers a plurality of users participating in a discourse related to the customer. The user type detector identifies a type of a user from the plurality of users as either real or fake and the attack identifier identifies an attack according to the type of the users.

[0008] Additionally, in accordance with an example, an attack is a negative campaign that include one or more fake users and is orchestrated by one or more real users.

10 [0009] Moreover, in accordance with an example, the system also includes a discourse analyzer to identify one or more key elements in the discourse and discover the plurality of users participating in conversations containing one of the key elements.

[0010] Furthermore, in accordance with an example, the user type detector also includes an activity analyzer to analyze an activity of each user and provide indications regarding the
15 type of the user according to the activity.

[0011] Additionally, in accordance with an example, the user type detector also comprises a language analyzer to analyze a language of each user and provide indications regarding the type of the user according to the language.

[0012] Furthermore, in accordance with an example, the user type detector also includes a
20 link analyzer to analyze a structure of a social network of each user and provide indications regarding the type of the user according to the structure of the social network.

Still further, in accordance with an example, the user type detector also includes a visual content analyzer to analyze a category of visual content, related to each user, the category of content is genuine, modified or generated and provide indications regarding the type of
25 the user according to the detected type.

[0013] Still further, in accordance with an example, the user type detector also includes a fake user detector to define whether a fake user is a bot or an avatar in a network of fake users.

[0014] Additionally, in accordance with an example, the user type detector also includes a unified entity creator to create for each user a unified entity representing the user in a plurality of networks.

[0015] Moreover, in accordance with an example, the attack identifier identifies an attack according to the number of users having a type fake.

[0016] There is provided in accordance with an aspect of the present disclosure a method for identifying attacks against a customer in a social network. The method includes discovering users participating in a discourse related to the customer, identifying a type of each user and identifying an attack according to the type of the users.

[0017] Additionally, in accordance with an example, an attack is a negative campaign that includes one or more fake users and is orchestrated by one or more real users.

[0018] Moreover, in accordance with an example, the method also includes identifying one or more key elements in the discourse and discovering the plurality of users participating in conversations containing one of the key elements.

[0019] Furthermore, in accordance with an example, the identifying includes analyzing an activity of each user and providing indications regarding the type of the user according to the activity.

[0020] Additionally, in accordance with an example, the identifying also includes analyzing a language of each user and providing indications regarding the type of the user according to the language.

[0021] Moreover, in accordance with an example, the identifying also includes analyzing a structure of a network of each user and providing indications regarding the type of the user according to the structure of the network.

[0022] Furthermore, in accordance with an example, the identifying also includes detecting a category of visual content related to each user, the category is genuine, modified or generated and providing an indication regarding the type of the user according to the detected type.

5 [0023] Still further, in accordance with an example, the identifying also includes defining whether a fake user is either a bot or an avatar.

[0024] Still further, in accordance with an example, the identifying also includes creating for each user a unified entity representing the user in a plurality of networks.

10 [0025] Moreover, in accordance with an example, the identifying also includes considering the unified entity.

[0026] There is provided in accordance with an aspect of the present disclosure a method for identifying a fake user in a social network. The method includes evaluating an authenticity of a user by applying a set of tests, each test in the set includes one or more rules on one or more parameters associated with the user, each test creating a weighted score. The
15 method also includes calculating a weighted average score from the weighted scores and determining if a user is fake according to the weighted average score compared to a configured threshold.

[0027] Additionally, in accordance with an example, a test is an artificial neural network (ANN), the rules are the structure of the ANN and the weighted score is the output of the
20 ANN.

[0028] Moreover, in accordance with an example, the set of tests includes analyzing the language of the user and extracting a set of characteristics, building an expected profile from the characteristics, comparing the expected profile with an actual profile provided by the user and creating a language score reflecting the degree of variation between the expected profile
25 and the actual profile.

[0029] Additionally, the set of tests includes detecting a category of visual content related to each user, the category is genuine, modified or generated, and providing an indication regarding the type of the user according to the detected type.

[0030] Furthermore, in accordance with an example, the detecting a category of visual content comprises using a generative adversarial network (GAN) to detect generated visual content.