Confianza y Reputación en Sistemas Multi-Agente

Jose Aguilar

CEMISID, Facultad de Ingeniería

Universidad de los Andes

Mérida, Venezuela

aguilar@ula.ve

La Confianza en los SMA

"La **confianza** es la probabilidad subjetiva por la que un individuo A espera que otro individuo B realice una determinada acción de la que el bienestar de A depende"

Gambetta 90

- RAE
  - ***Confianza*****:** *esperanza firme que se tiene de alguien o algo*
  - ***Reputación*****:** *prestigio o estima que son tenidos de alguien o algo*

- Falta de consenso semántico
  - Confianza ≠ Reputación
  - Confiar (Trust) ≠ Reputación ≠ Confianza (Confidence) ≠ Confiabilidad (Reliability) ≠ Voluntad (Willingness) ≠ …

La Confianza y Reputación en los SMA

- ¿Por qué hablar de ellos en los SMA?
- ¿ Qué podemos extraer de las sociedades humanas, de animales?
- ¿ Qué lo hace distintos de SMA que no consideren ese aspecto?

  => Sistemas orientados a servicios

- ¿ Podrian asociarse a aspectos de auto-regulación, mecanismos de auto-reparación, etc.?

La Confianza en los SMA

- El agente Ag1 **cree** que el agente Ag2 puede y desea hacer una acción dada;
- El agente Ag1 **tiene la meta** de que el agente Ag2 haga esa acción; y
- El agente Ag1 **confía** en el agente Ag2, absteniéndose de ejecutar la acción delegada y **coordinando su comportamiento con el comportamiento esperado** del agente Ag2.

La Confianza en los SMA

- Dificultad de modelización
  - Medida de la estima o del prestigio
  - Subjetividad
    - ¿Qué es algo bueno? ¿Para quién?

- Dificultad para el diseño
  - Medidas cuantitativas o cualitativas…
    - $t_{A \rightarrow B} = 0.7$ | $t_{B \rightarrow A} = 0.3$ | $t_{B \rightarrow C}$ = “bueno” | $t_{B \rightarrow A} = k$

- Distintos campos confluyen: informática (modelos computacionales), sociología (redes sociales), psicología (estados mentales), economía (funciones de utilidad, toma de decisiones, etc.), …

Modelos de confianza

- ¿Cómo lograr sistemas estables usando la confianza?

1. Empatía o reciprocidad entre los agentes a partir de factores como la disposición, la fiabilidad, etc.
   - Se caracterizan (dinámicamente) a través de **Modelos de Confianza**
   - La confianza será una medida para **decidir la interacción** con otro agente (necesaria pero no suficiente)
   - Interacciones **directas e indirectas**
   - Usa información **interna** (p.e., experiencia pasada) o **externa** (p.e., opiniones de terceros)

2. ~~Diseño de protocolos y mecanismos para las interacciones (reglas de encuentro)~~

Modelos de confianza

- Dos formas de conceptualizar la confianza en sistemas multi-agente
  - *Nivel individual*: creencia sobre el comportamiento futuro de otro agente en una situación determinada

    → de interés para el agente
  - *Nivel de sistema*: usada para forzar a los agentes a actuar de forma benevolente (comportamiento emergente)

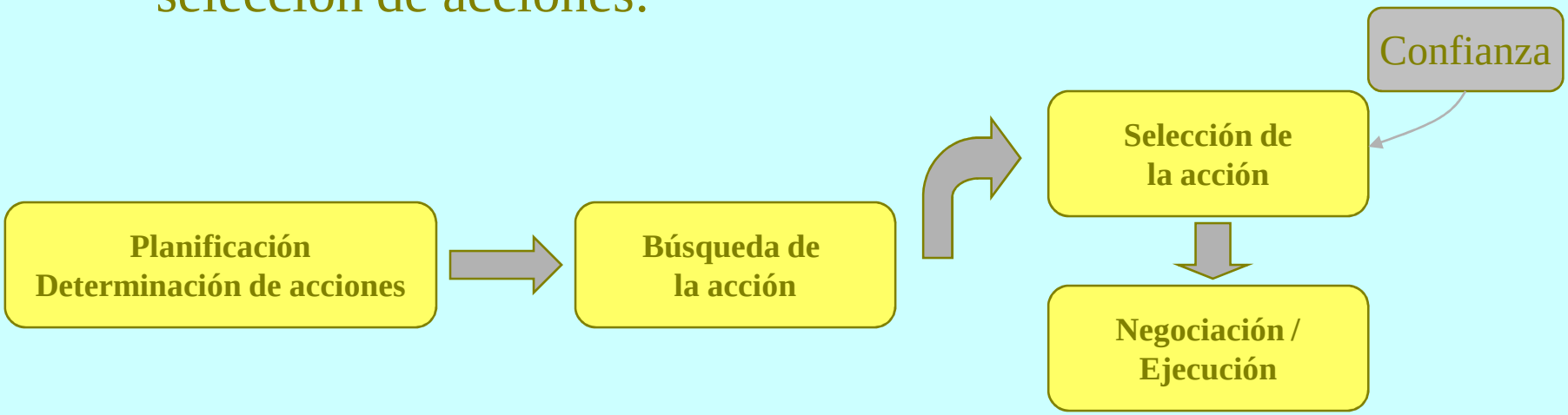
    → de interés para el sistema

- Conceptos implícitos:

  honestidad, benevolencia, malicia, incompetencia, altruismo → difíciles de distinguir!!

Modelo de Confianza en SMA

- Sirven para estimar la confianza en un agente/situación
- Predicen el comportamiento en una hipotética interacción
- *confianza* = $f(\text{experiencia}_{local}, \text{reputación}, ...)$
- Confianza ≠ Confiar
- La confianza se usa como un factor en procesos de selección de acciones:



Modelos de confianza

Modelos Centralizados vs. Distribuidos

- Centralizados
  - Mejor gestión de la información
  - Presunción de autoridad (¿debemos confiar en ella?)
  - Menos flexible / más dependiente
- Distribuidos
  - Permiten subjetividad
  - Mecanismo adaptativo y autónomo de control social → normas sociales emergentes (sociedades humanas)
  - Antropomorfismo
  - Más robusto

Modelos centralizados y distribuidos → ¡¡¡compatibles!!!

Modelos de confianza

- Problemas
  - Ajuste de la confianza tras una experiencia directa
  - Ajuste de la confianza tras una experiencia indirecta (p.ej. en una organización o grupo)
  - Cómo agregar opiniones de otros a las experiencias locales propias
  - Cómo ajustar la confianza en agentes que han servido como fuentes de reputación
  - Exploración vs. Explotación

Modelos de confianza

- Clasificación de los modelos de confianza
  - Modelos basados en Aprendizaje y Evolución
  - Modelos basados en Reputación
  - Modelos Socio-Cognitivos
  - Modelos para Organizaciones
  - Alineación semántica

Modelos de confianza

EVOLUCIÓN

- Confianza emerge como resultado de la evolución de estrategias en varias interacciones [Wu,Sun]
  - Mayor ganancia a largo plazo para los agentes “buenos” sacrificando parte de su utilidad a corto plazo (*Dilema del prisionero)*
  - Umbral en el número de interacciones (característico para el cálculo)
  - Variabilidad de los resultados

- No siempre escenarios cooperativos

- No son ideales
  - Malevolencia
  - Incompetencia
  - Desconocimiento, …

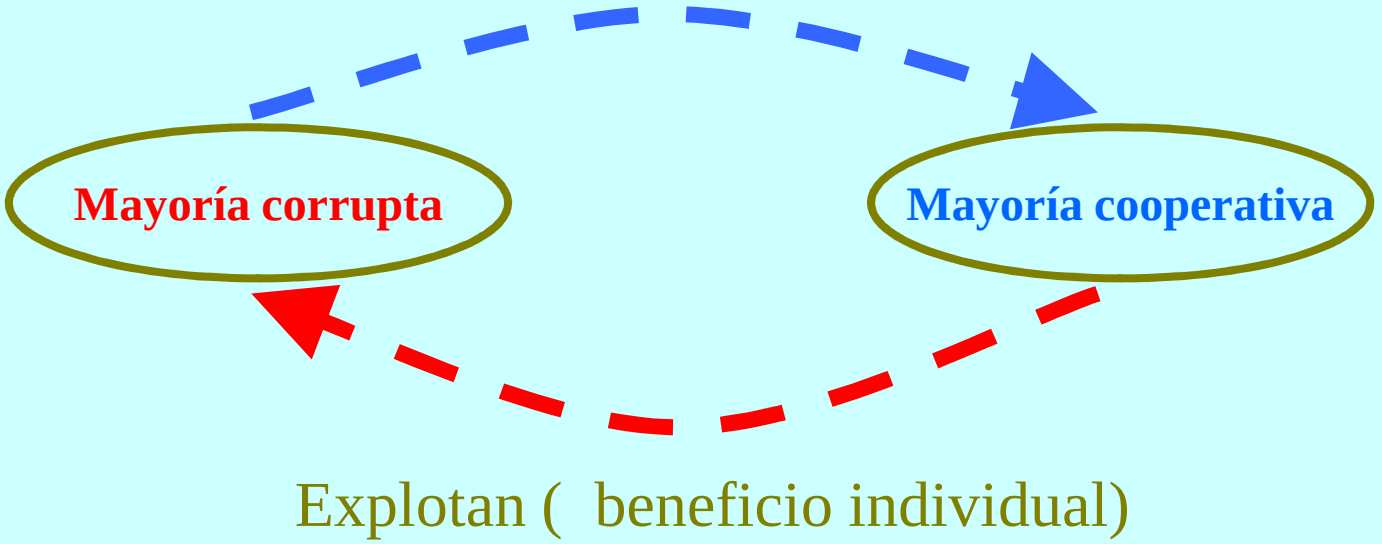
- Utilidad vs. Confianza (equilibrio)

Modelos de confianza basados en aprendizaje

- Confianza emerge del aprendizaje [Birk]
  - Dar más peso a objetivos sociales vs. objetivos individuales
  - Metáfora de la sociedad (inversión)
  - Ciclo de actuación

Equilibrio final

Ciclo de Birk



Modelos de confianza basados en Reputación

- Opinión o vista de alguien acerca de algo

  Confianza ≠ Reputación

- Reputación es un medio para la Confianza (ej. eBay)
- Líneas de investigación en Reputación:
  - Métodos para recoger *ratings* acerca de lo "bueno" que es el agente
  - Métodos de razonamiento para la agregación de esa información
  - Mecanismos de actualización (nuevos pesos)
- Recuperación y agregación con Redes Sociales
  - Representación mediante grafos (relaciones)
  - Los *ratings* "fluyen" entre los nodos por los arcos
  - Honestidad y Altruismo [Schillo et al.]
  - Posibilidad para los agentes de adquirir información de otros

Modelos de confianza basados en Reputación

- Agregación
  - Combinación de valores de reputación que llegan a un agente
  - Ejemplo: eBay (+1, -1). Comentarios?
    - Valoraciones detalladas del vendedor con cuatro categorías diferentes.
    - Los compradores se les pidió que calificaran el vendedor en cada una de estas categorías con una puntuación de una a cinco estrellas, siendo cinco la calificación más alta y uno más bajo.
    - Estas calificaciones son anónimas.
    - Las listas de los vendedores con una calificación de 4.3 o más en cualquiera de las cuatro categorías de calificación aparecen antes en los resultados de búsqueda.
    - Tiendas de alimentación están obligados a tener las puntuaciones de cada categoría por encima de 4,5.

Modelos de confianza basados en Reputación

- Agregación
  - Modelos basados en teorías más sólidas:
    - Teoría de Demspter-Shafer → Funciones de creencia - tres tipos de valores:
      - *Trustworthy* (digno de confianza)
      - *Untrustworthy*
      - *Uncertain*

Modelos de confianza basados en Reputación

- Concepto de *witness*
  - Testigos o fuentes de reputación
- Problema:
  - Asunción de que nunca mienten
  - Cómo elegir las fuentes de reputación
  - Todo intercambio de información implica riesgo de (casual o intencionada) → A más intercambio mayor riesgo de imprecisión
- Soluciones:
  - *honestidad* y *altruismo* [Schillo]
  - probabilidad de mentira de los testigos [Sen y Dutta]

Modelos de confianza basados en Reputación

- 3 dimensiones de la reputación [Sabater y Sierra] :
  - Dimensión individual
    - Reputación como suma ponderada de impresiones subjetivas que derivan de interacciones directas
  - Dimensión social
    - Impresión del grupo de *a1* acerca de *a2*
    - Impresión del grupo de *a1* acerca del grupo de *a2*
    - Impresión de los agentes del grupo de *a2* sobre su compañero
  - Dimensión ontológica

Modelos de confianza basados en Reputación

- Mecanismo de confianza basado en reputación:
  - Más valor a experiencias propias
  - Cuanto mayor es el intercambio de información mayor es la probabilidad de hacer un razonamiento inexacto (incompetencia – malevolencia)
  - Buscar un equilibrio entre
    - Razonamiento local → experiencia propia
    - Información externa → reputación

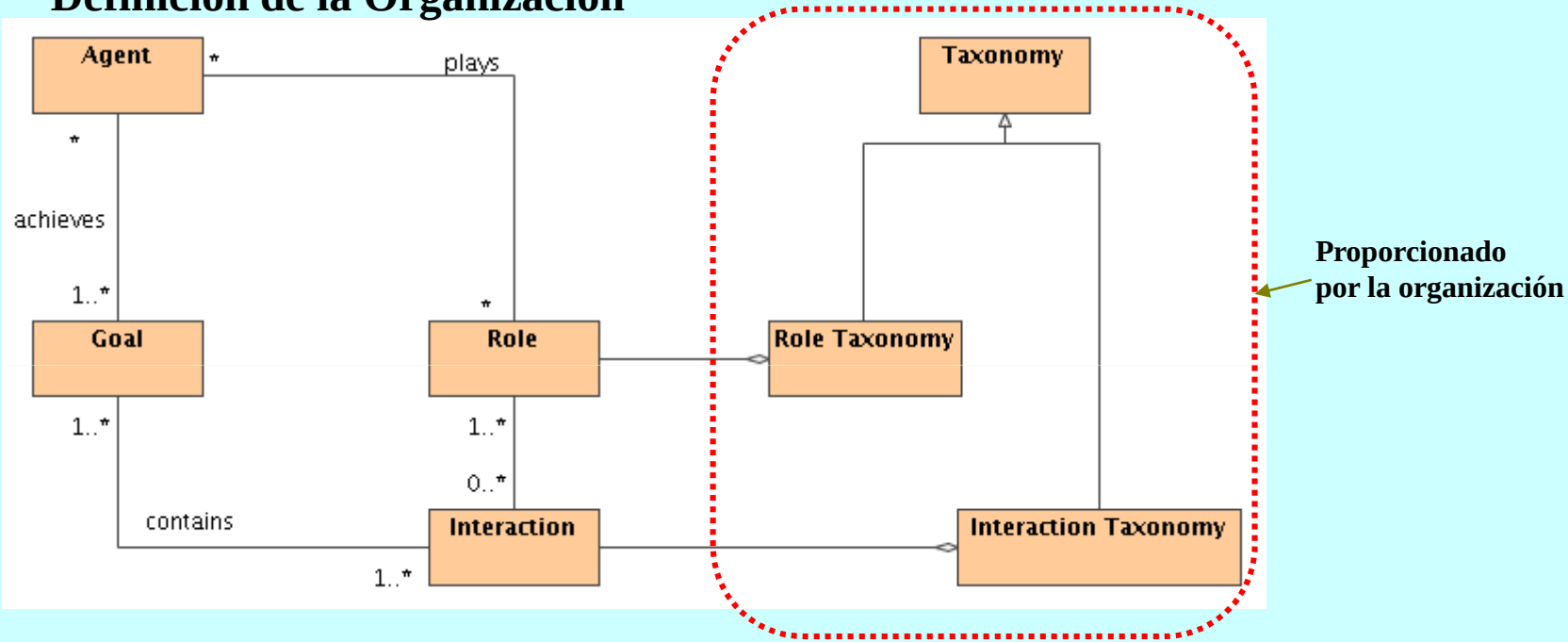
Confianza

Modelos de confianza para Organizaciones

- Modelos de confianza orientados a organizaciones de agentes
  - Organización → SMAs + restricciones
    - Roles, interacciones, normas, …
  - Aprovechan la información de la organización para mejorar modelos de confianza
    - Similitud de roles/interacciones
    - Argumentación con normas

Modelos de confianza para organizaciones

Definición de la Organización



Problema de decisión del agente

- Los agentes deben cumplir objetivos
- Los agentes deben realizar diversas interacciones para cumplir estos objetivos
- De entre los agentes capacitados (por su rol) para realizar las interacciones, ¿a cuál de ellos elegir?

Modelos de confianza para Organizaciones

- Los agentes mantienen una **Local Interaction Table (LIT):**
  - Estructura de datos que compila los resultados de las interacciones pasadas
  - Indexados por claves del tipo agente/rol/interaction (**<X,Y,Z>**)
  - Con valores asociados de **confidence** y **reliability**

$\langle X, Y, Z \rangle$	$c_{A \rightarrow \langle X,Y,Z \rangle}$	$r_{A \rightarrow \langle X,Y,Z \rangle}$
$\langle a_2, r_5, i_1 \rangle$	0.5	0.3
$\langle a_4, r_1, i_2 \rangle$	0.7	0.8
$\langle a_2, r_3, i_1 \rangle$	0.9	0.5
⋮	⋮	⋮
$\langle a_9, r_2, i_5 \rangle$	0.4	0.7

  - **<X,Y,Z>**: unidad atómica "agente X jugando el rol Y en la interación Z"
  - **Confidence (cA→ <X,Y,Z>)**: de la propia experiencia del agente
  - **Reliability (rA→ <X,Y,Z>)**: mide cómo de seguro está un agente acerca de su propio valor de *confidence* en la situación <X,Y,Z>

Modelos de confianza para organizaciones

- **Enfoque básico**:
  - *"… un agente A jugando el rol Estudiante necesita encontrar un compañero para un tipo de práctica en una asignatura específica"*
  - El *estudiante* mirará en su LIT las interacciones pasadas haciendo ese tipo de prácticas (**I**) con otros *estudiantes* (**R**)
  - El agente A seleccionará a aquel agente con una mayor *confidence* en su LIT poniendo atención únicamente en aquellas entradas **<X,R,I>**, donde X puede ser cualquier otro agente

Modelos de confianza para organizaciones

- Problema: la LIT puede ser incompleta
  - **No hay entradas** para la situación <X,Y,Z>
  - Existe **baja precisión** en los valores para las entradas <X,Y,Z>

- Hipótesis: los agentes se comportan de manera similar cuando juegan roles similares o actúan en interacciones similares
  - Los valores de **confidence** pueden ser aproximados a partir de otras entradas en la LIT del agente en lugar de preguntar a fuentes de reputación
  - Experiencias con un agente en roles/interacciones **similares** pueden ser usadas para aproximar mejor valores de **confidence**
  - Las **similitudes** pueden ser obtenidas a partir de las taxonomías de roles e interacciones que facilita la organización

Modelos de confianza

- **Basic trust model** aplicado a los datos de la LIT:

$$t_{A\rightarrow\langle B,R,I\rangle} = \frac{\sum_{\langle X,Y,Z\rangle \in LIT_A} c_{A\rightarrow\langle X,Y,Z\rangle} \cdot w_{A\rightarrow\langle X,Y,Z\rangle}}{\sum_{\langle X,Y,Z\rangle \in LIT_A} w_{A\rightarrow\langle X,Y,Z\rangle}}$$

- Sacamos partido a la estructura de la organización (**inference trust model**):

$$w_{A\rightarrow<X,Y,Z>} = r_{A\rightarrow<X,Y,Z>} \cdot sim(<X,Y,Z>,<B,R,I>)$$

- Cálculo de la **Similitud**:

$$sim(\langle X,Y,Z\rangle,\langle B,R,I\rangle) = \begin{cases} \beta \cdot sim_R(R,Y) + \gamma \cdot sim_I(I,Z), & if \quad B = X \\ 0, & otherwise \end{cases}$$

- **Actualización de confidence**:

$$c_{A\rightarrow\langle X,Y,Z\rangle} = \epsilon \cdot c'_{A\rightarrow\langle X,Y,Z\rangle} + (1-\epsilon) \cdot g_{\langle X,Y,Z\rangle}$$

- **Actualización de la reliability**: basado en el número de interacciones y en su variabilidad

Modelos de confianza Comerciales

- Dominios comerciales o redes sociales
- Ejemplo: eBay
  - Usado como ayuda al usuario
  - Valoración de opiniones distribuida
  - Visión de valoraciones unívoca
  - Naturaleza de las valoraciones:
    - Suma de opiniones numéricas individuales: +1,0,-1
    - Comentarios textuales
  - Visión global de la reputación de los usuarios

Modelos de confianza Comerciales

- Trip Advisor (www.tripadvisor.es)
  - Opiniones
  - Votaciones
  - Rankings
    - Positivos (preferencias)
    - Negativos (características a evitar)
    - Índice de popularidad de TripAdvisor le incorpora Contenido Traveler para determinar la satisfacción del viajero

Modelos de confianza Comerciales

- ¿Cómo saber si un hotel en particular es adecuado para usted? El precio de la habitación puede ajustarse a su presupuesto, pero las habitaciones están limpias? Es el lugar seguro? ¿Cómo es el servicio?
- Índice de popularidad para ayudar rápidamente a responder esas preguntas.

  Resultados que influyen en ranking.
  Constantemente incorpora nueva información.
  Global: Refleja las opiniones de todo el mundo.
  Imparcial: Basado en la valoración de todos
- Esto asegura que siempre se busca en e los mejores hoteles.

Modelos de confianza Comerciales

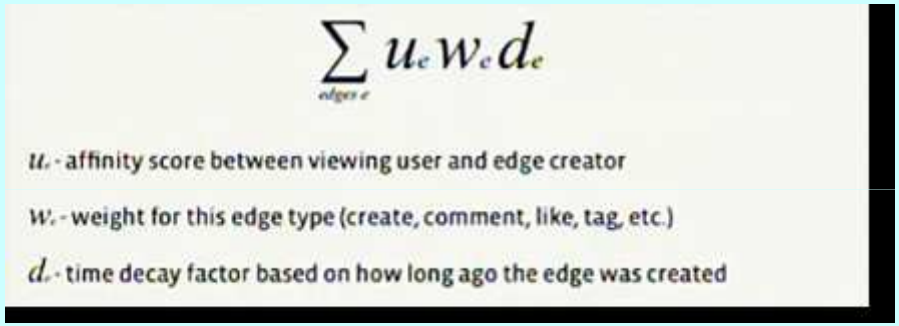
- Dominios comerciales o redes sociales
- Ejemplo: eConcozco, Friendster, Facebook, . . .
- Aplicable a todo tipo de interacciones
  - Intercambio de datos
    - Redes P2P
    - Descarga de ficheros
      - Usan valoraciones como indicación al usuario
      - Opiniones
      - Ratings

        1..10

        “muy bueno”

Modelos de confianza Comerciales

Facebook



- Hay tres principales factores de ranking :
  Afinidad: Esto está directamente relacionado a los gustos, las interacciones y sus relaciones.
  Peso: Piense en acciones que tienen un mayor esfuerzo
  Relevancia y actualidad

Modelos de confianza Comerciales Facebook

¿Qué Rank significa para el futuro?

Una página con muchos fans y poca o ninguna interacción recibiría la puntuación más baja.

¿Cómo ir Arriba?

Utilice fotos y vídeos a menudo.

Fomentar la interacción sin ser agresivo.

Haga preguntas a sus fans y amigos no pueden dejar de responder.

Crear una encuesta…