

Oriëntatie Kunstmatige Intelligentie

Representatie en Geheugen

Niels Taatgen

Representatie en Geheugen

- Kunstmatige Intelligentie: bestudeer **Intelligentie** met **formele systemen**
- Formeel systeem: verzameling afspraken
- Vraag: Wat is de beste afspraak om menselijke kennis op te slaan?
- Methode: Probeer te ontdekken hoe mensen hun kennis representeren: bestudeer Geheugen

Inhoud

■ Drie benaderingen

- Informatica (3.1, 3.2, 3.5, 5.3)
- Cognitieve Psychologie (5.1-5.4)
- Neurowetenschap (2)

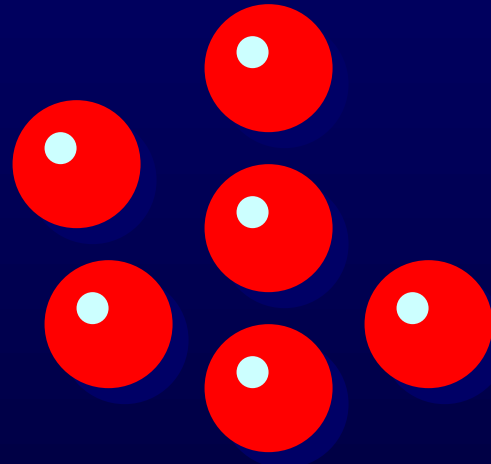
■ Komen samen in

- Neurale netwerken (3.5, 5.3, 7.5-7.6)
- ACT-R (8.2-8.3)

Zes

|

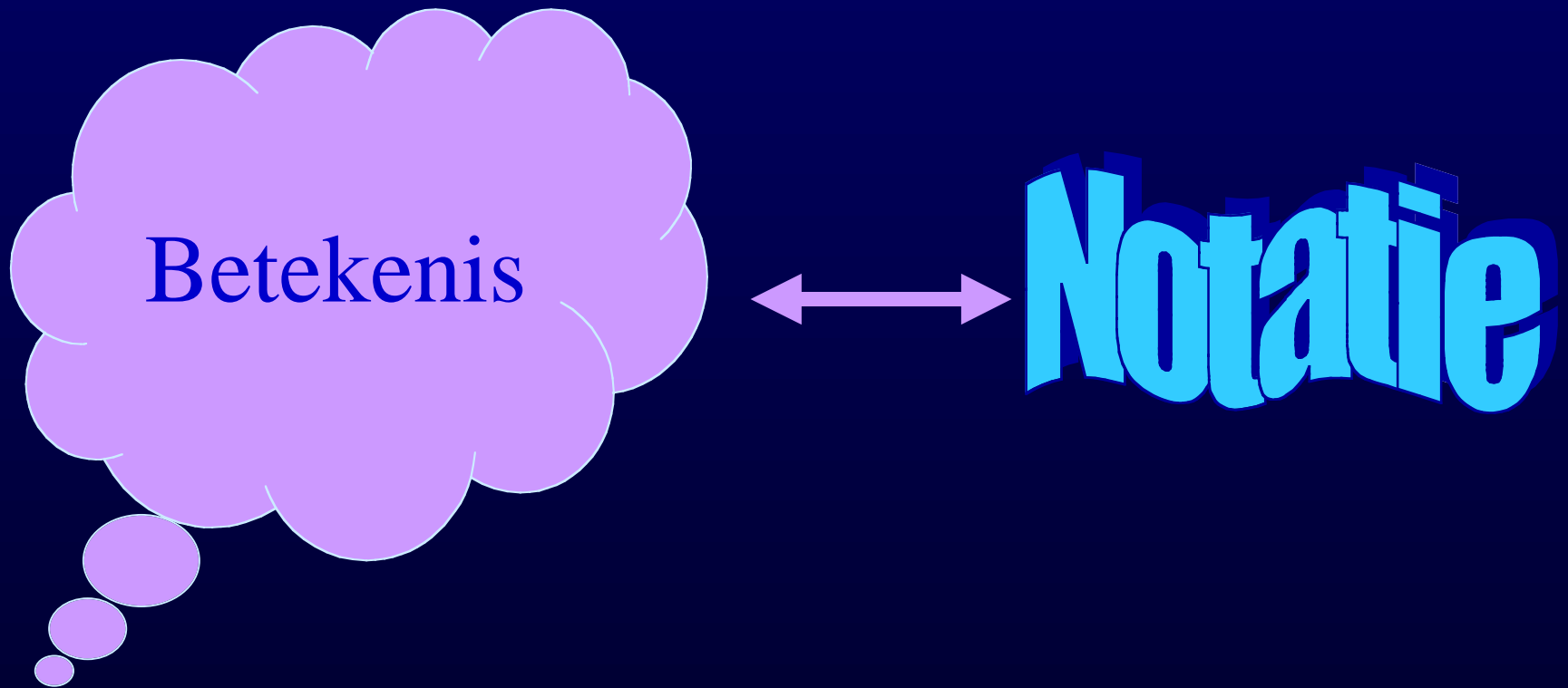
6



VI

110

Representatie



Representatie

Eh..
Zestien plus
vijfentwintig is..
eenenveertig



$$\begin{array}{r} 16 \\ 25 \\ \hline 41 \end{array} +$$

Lichaam-geest probleem

Representatie van Handelingen

per
he



*Dit is echter niet de
manier waarop mensen
kennis in hun hoofd
opslaan!*

he

vooruit

rechts(90)

einde herhaal

einde herhaal

pen_omhoog

Het probleem van representatie

- Wat is betekenis nu eigenlijk?
- Hoe representeren mensen feitenkennis?
- Hoe representeren mensen kennis over handelingen, d.w.z. kennis over hoe je iets moet doen?

Geheugen

- Behavioristen (begin vorige eeuw):
b.v. Pavlov

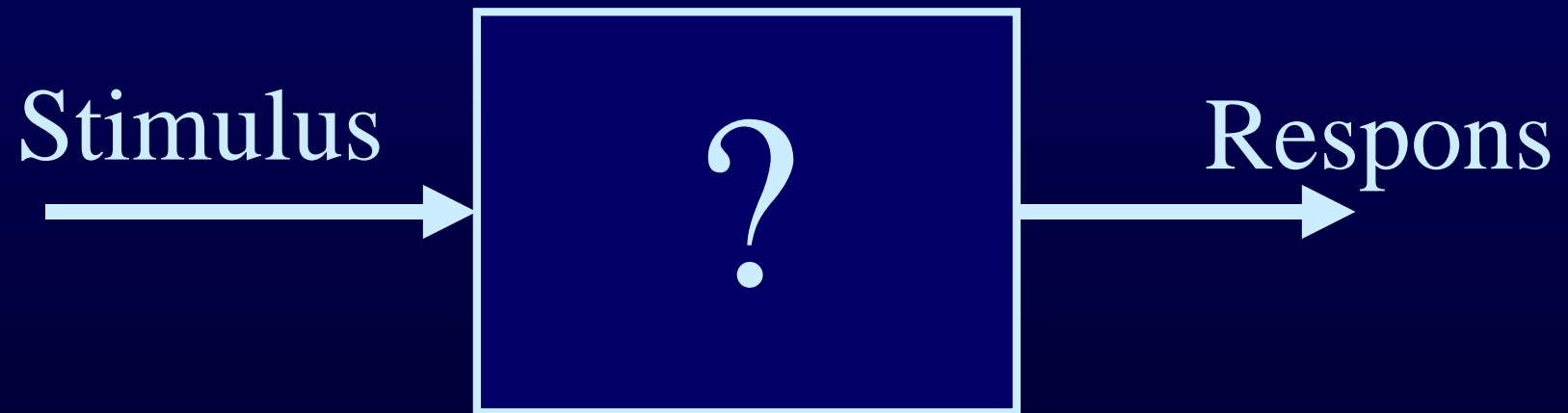


© 1997 The Learning Company, Inc.



Geheugen

- Behavioristen (begin vorige eeuw)

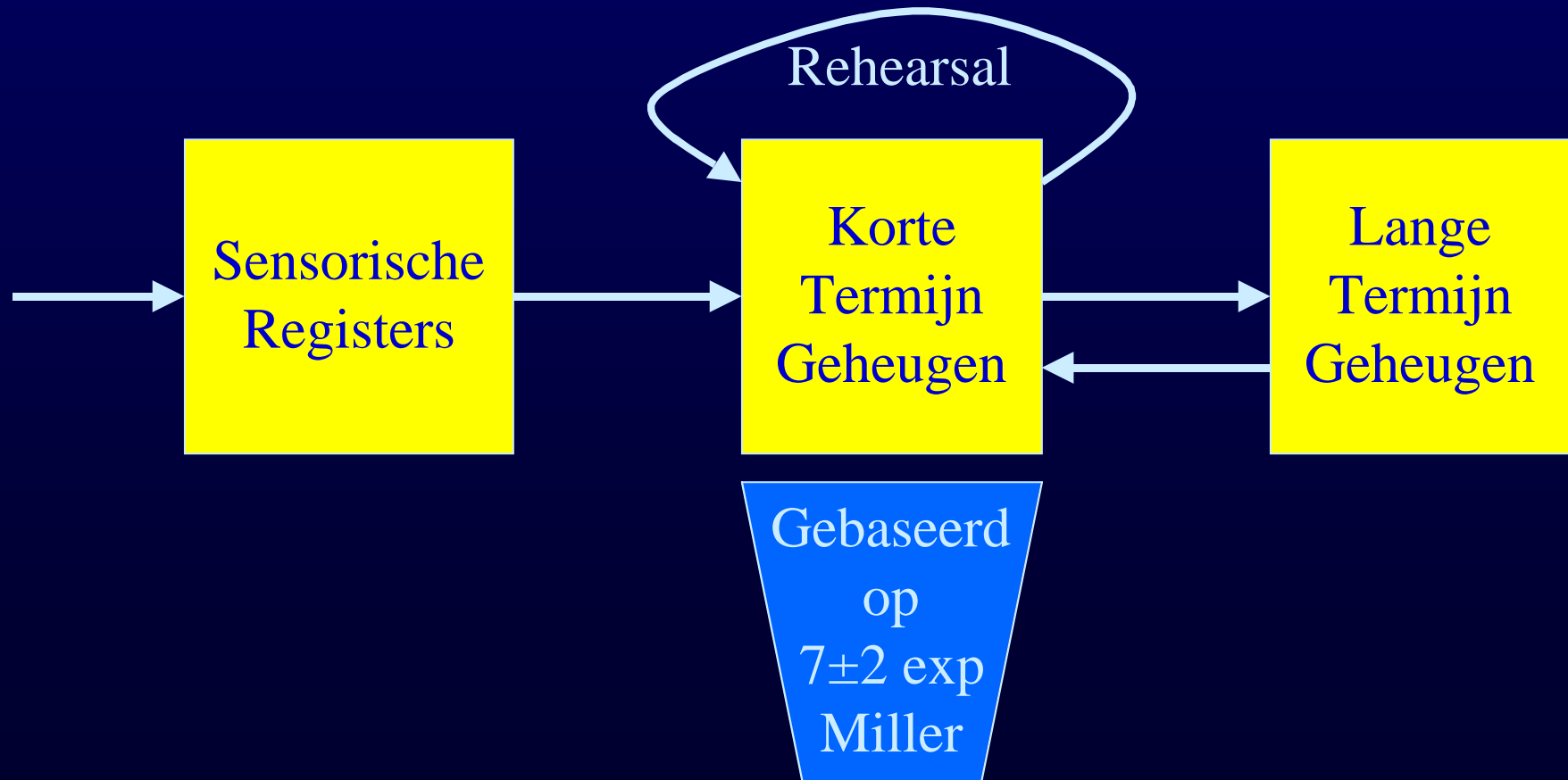


Geheugen

- Behavioristen (begin vorige eeuw)



Atkinson & Shiffrin (1968)



Een geheugenexperiment

- Je krijgt zometeen een aantal woorden te zien (het zijn Engelse woorden). Probeer deze zo goed mogelijk te onthouden.

(Niet opschrijven!!!!)

Avocado

Hexagon

Cinnamon

Broccoli

Electron

Flamingo

Aardvark

Boyhood

Crevice

Gazette

Anatomy

Bourbon

Nu komt de test

De bedoeling is dat je de woorden afmaakt die zometeen op het scherm verschijnen. Steek je vinger op als je het woord herkent (of denkt te herkennen)

AL__N_C

AN_ _ _MY

B _ _ CC _ _ I



BE__VI__



_OYH_O_



B__ZI_R

Einde van de test

- Sommige woorden stonden op de lijst, anderen niet
- In een “echt” experiment (met meer woorden), scoren proefpersonen:
 - 50% goed op woorden uit de lijst
 - 30% goed op woorden die niet op de lijst stonden

Nu komt nog een tweede test

Indien het woord dat op het scherm verschijnt in de lijst stond, steek dan je vinger op, anders niet.

Aardvark

Flamingo

Cockatoo

Ellipse



Gazette

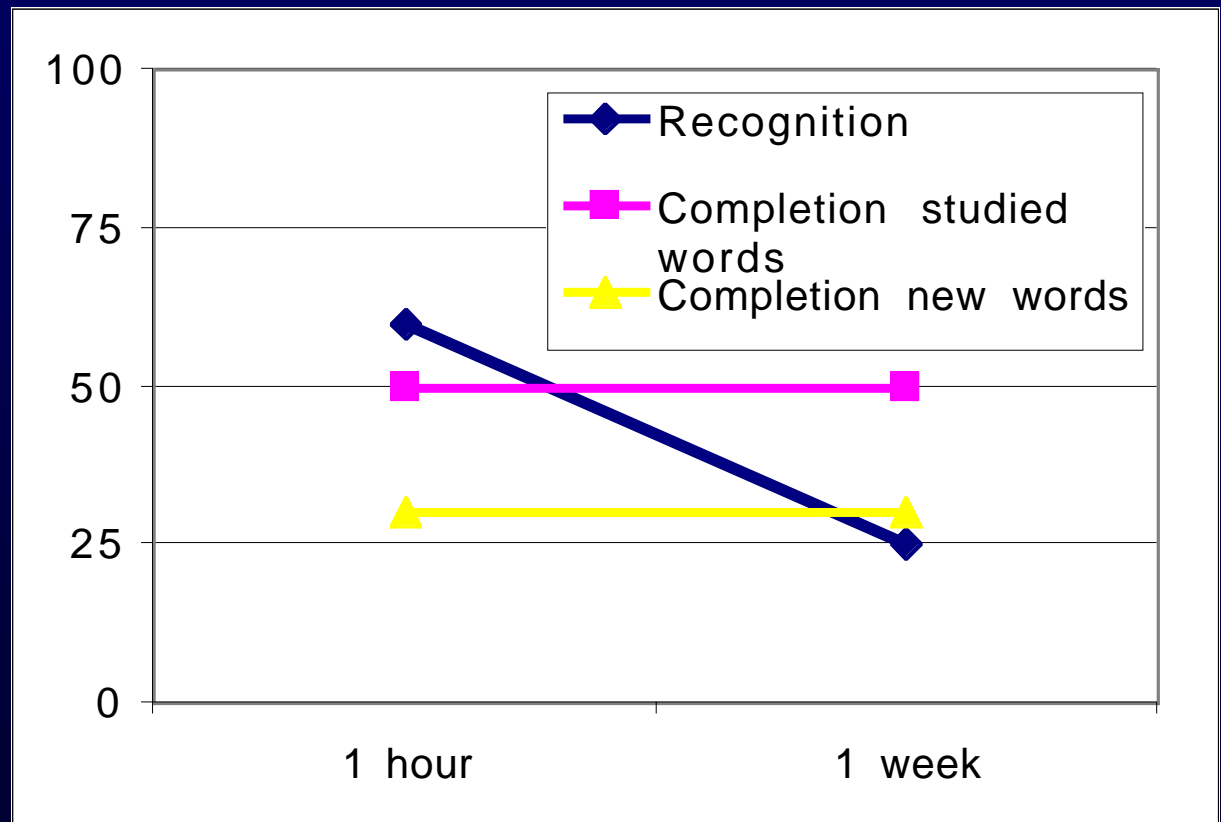
Airspace

Einde van de test

In een “echt” experiment hebben proefpersonen na een uur ongeveer 60% goed op deze test, waarbij inbegrepen is de gokkans.

Experiment van Tulving e.a.

96 woorden
Test na een
uur
Test na een
week



Uit een ander experiment

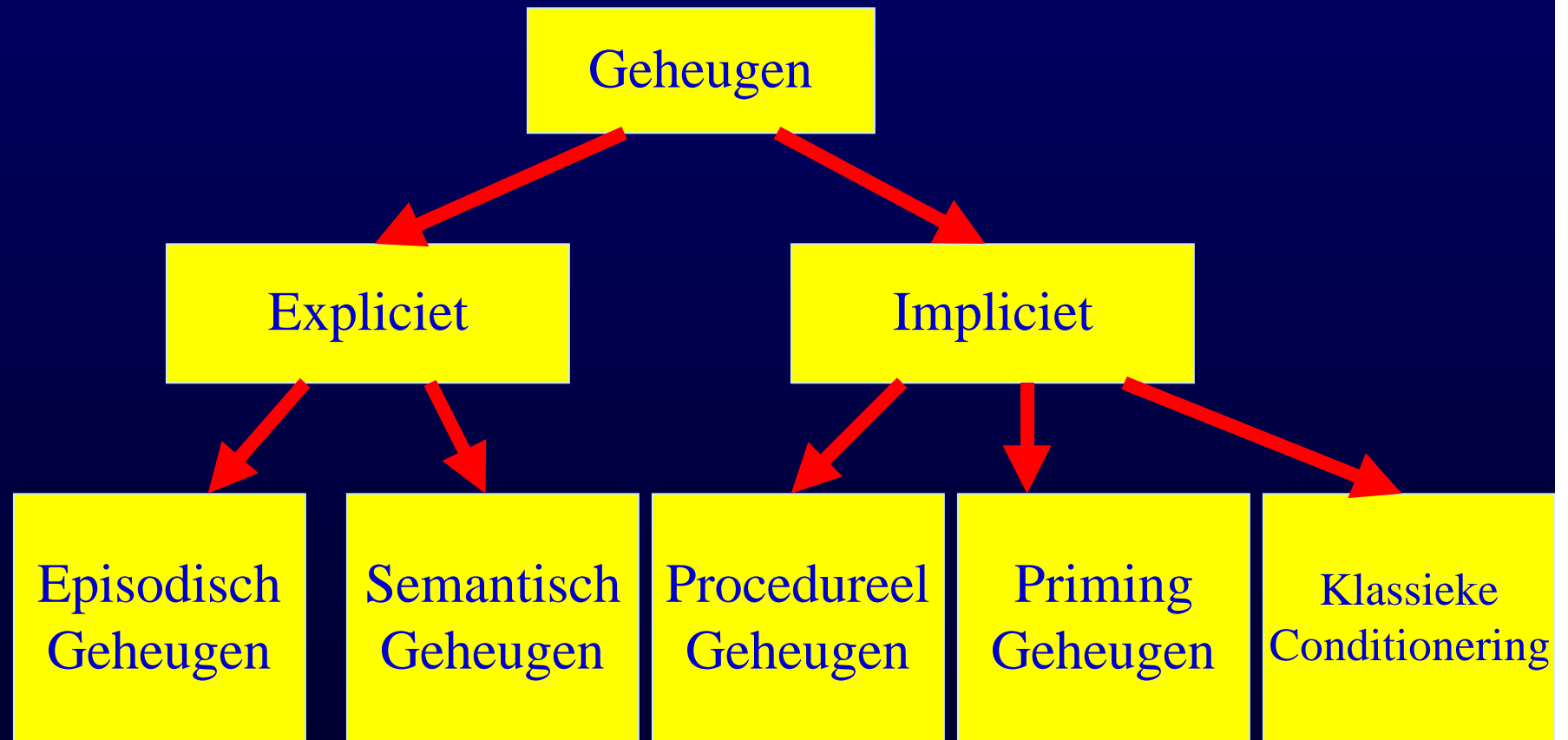
Als je hetzelfde experiment met amnesiepatiënten (d.w.z. geheugenverlies) doet blijkt:

- Amnesiepatiënten zijn heel slecht in herkenning
- Ze zijn echter net zo goed als proefpersonen in het afmaken van woorden.

Conclusie Tulving & co.

- Mensen hebben twee soorten lange-termijn geheugen
 - Expliciet, bewust geheugen (herkenningstaak)
 - Impliciet, onbewust geheugen (woorden afmaken)

Squire & Knowlton (1995)



Probleem van geheugenonderzoek

- Hoe ver moeten we gaan met het opsplitsen van geheugen?
- Hoe is kennis in elk van deze geheugensystemen gerepresenteerd?
- En hoe werken al deze geheugens samen?

Conclusie psychologisch onderzoek

- De psychologie levert een grote verzameling feiten over het geheugen, waarvoor we uiteindelijk een verklaring moeten bieden vanuit onze geheugentheorie
- De psychologie biedt betrekkelijk weinig theorie voor de vraag hoe kennis nu echt gerepresenteerd is

Geheugen is onderdeel van de hersenen

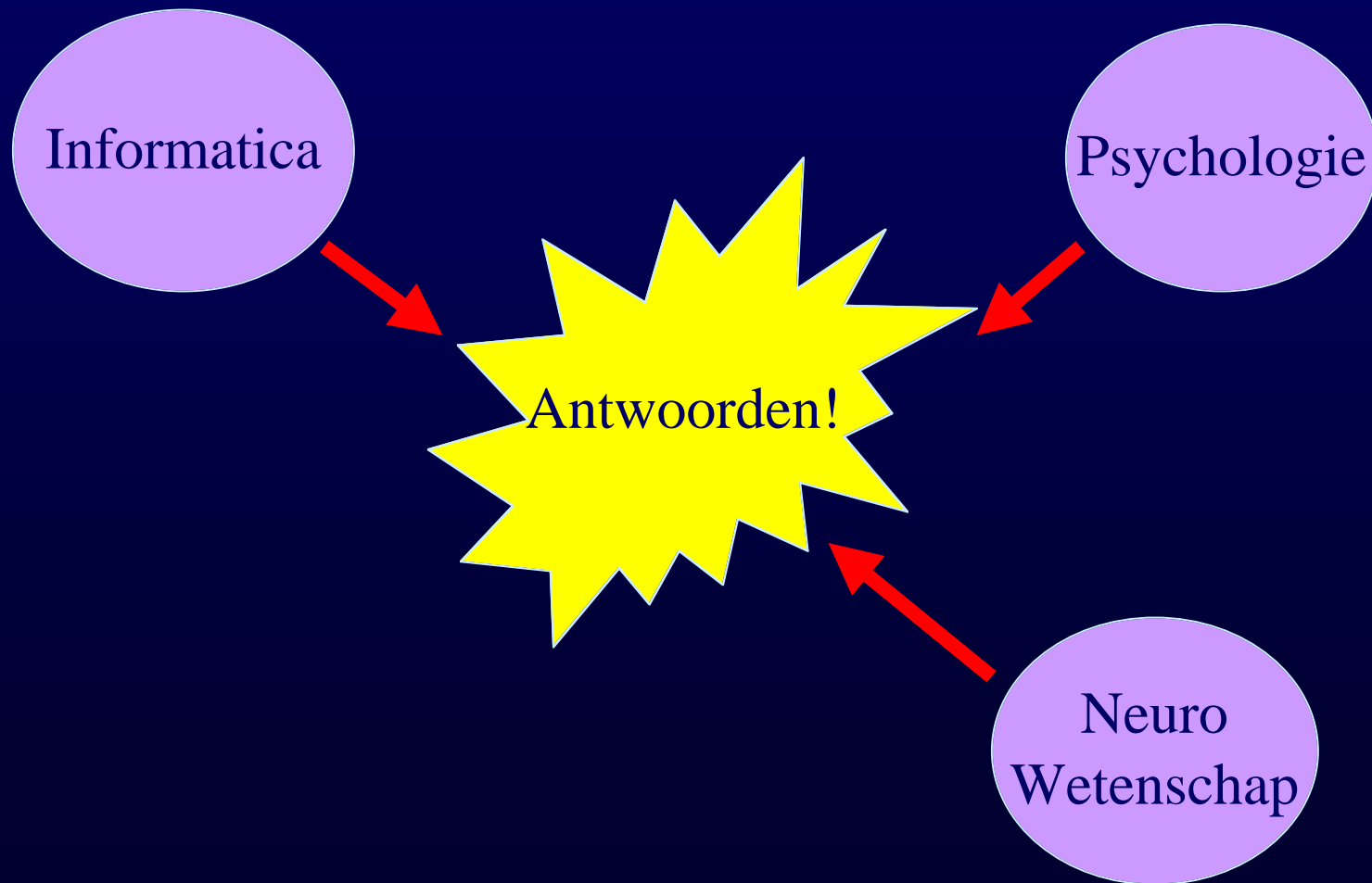
- Hersenen vormen een netwerk van miljarden neuronen
- Elke neuron is verbonden met honderden anderen
- Neuronen communiceren met elkaar door middel van chemische stoffen, neurotransmitters

Vragen

- Waar zit dan het geheugen?
- Hoe kunnen hersencellen (symbolische) kennis representeren?
- Hoe kunnen ze deze kennis leren?

(advertentie)

Hier begint de Kunstmatige Intelligentie

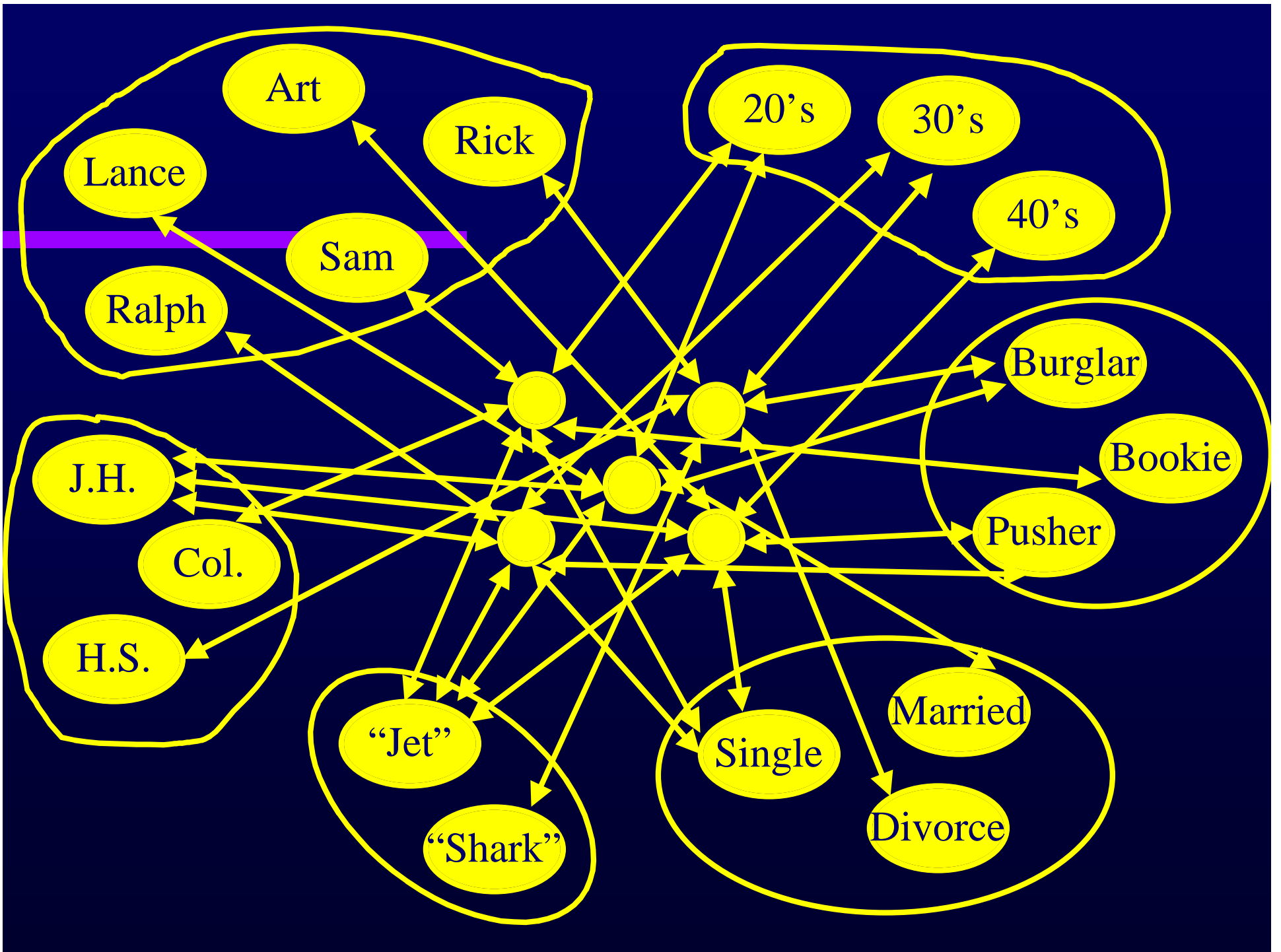


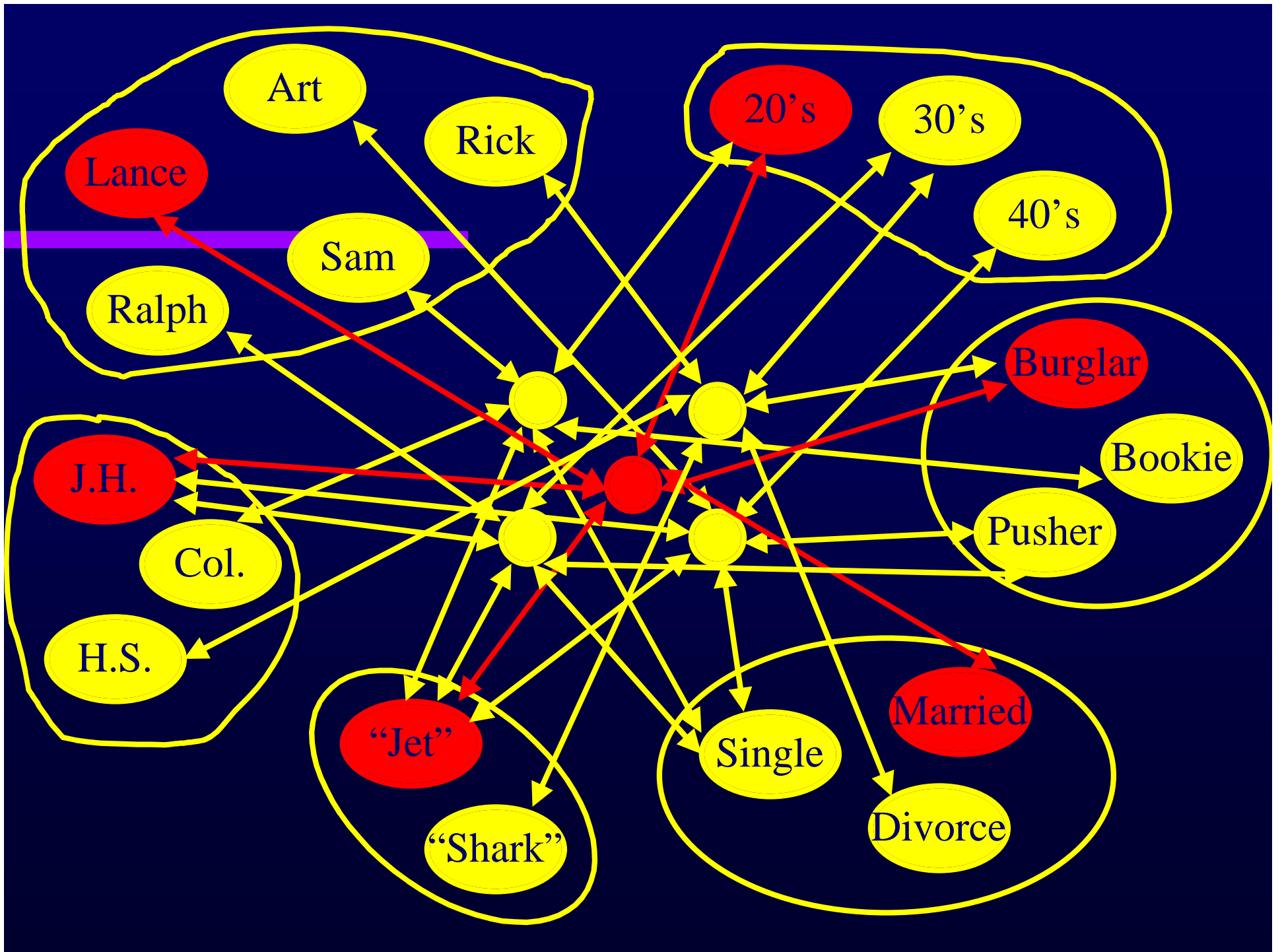
Neurale netwerken

- Neuronen hebben twee soorten verbindingen
 - Exciterend: als de ene cel actief is, wordt de cel die ermee verbonden is ook meer actief
 - Inhiberend: als de ene cel actief is, wordt de activiteit van een andere cel onderdrukt

The Jets and the Sharks

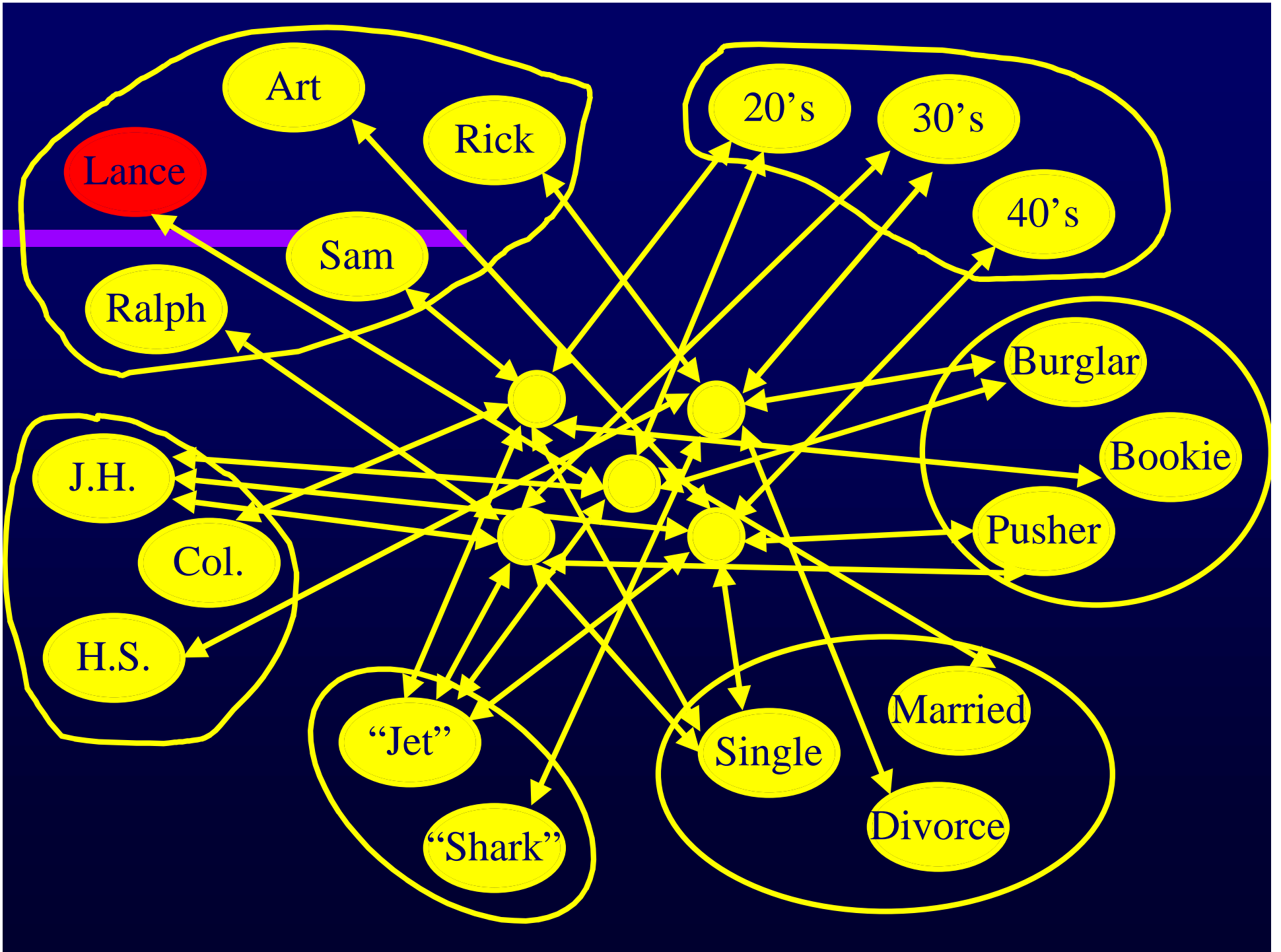
Name	Gang	Age	Edu	Married	Occup
Art	Jets	40's	J.H.	Sing.	Pusher
Sam	Jets	20's	Col.	Sing.	Bookie
Lance	Jets	20's	J.H.	Mar.	Burgler
Ralph	Jets	30's	J.H.	Sing.	Pusher
Rick	Shark	30's	H.S.	Divorc	Burgler

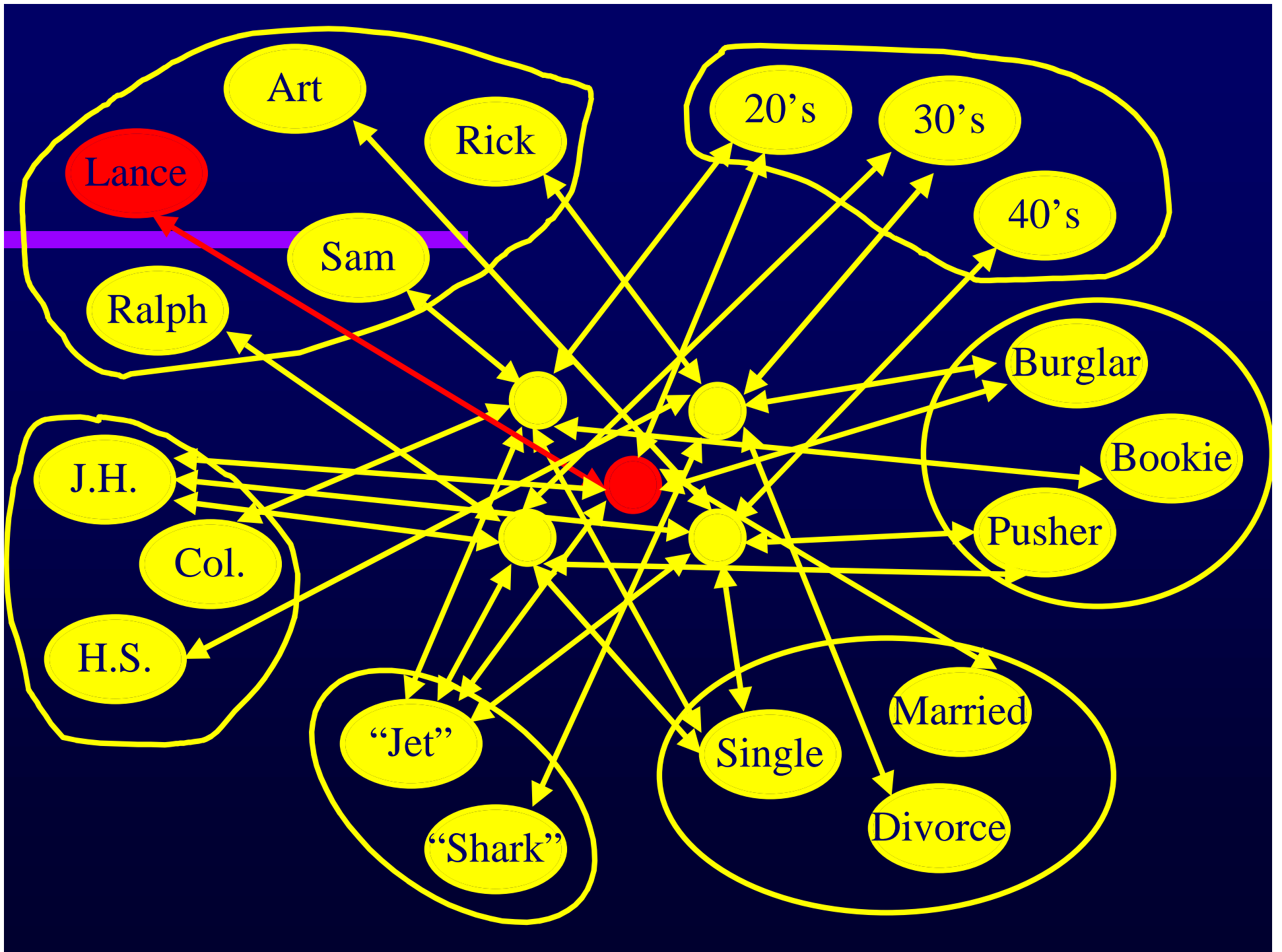




Vraag aan het netwerk

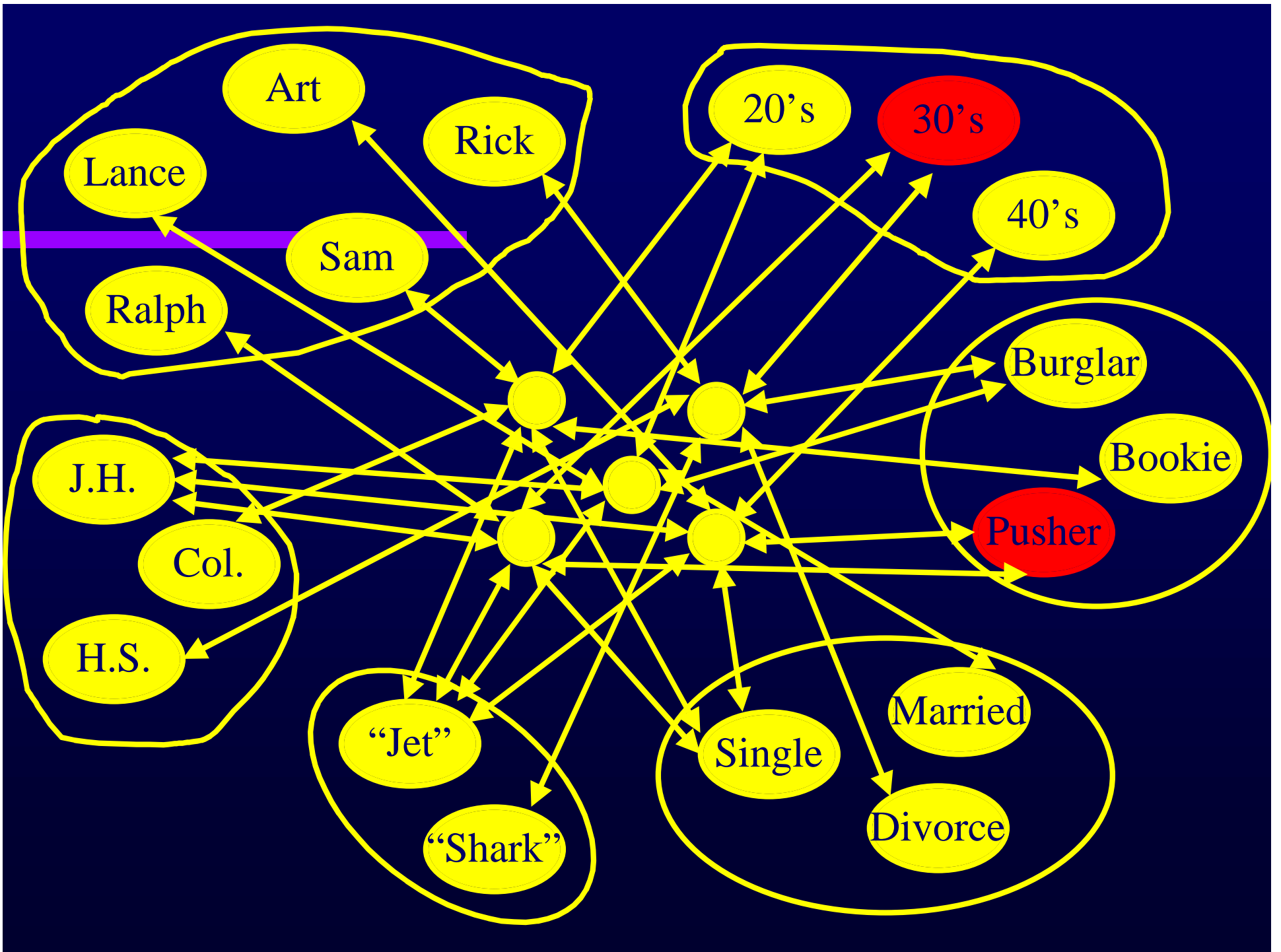
- Wat zijn de eigenschappen van Lance?

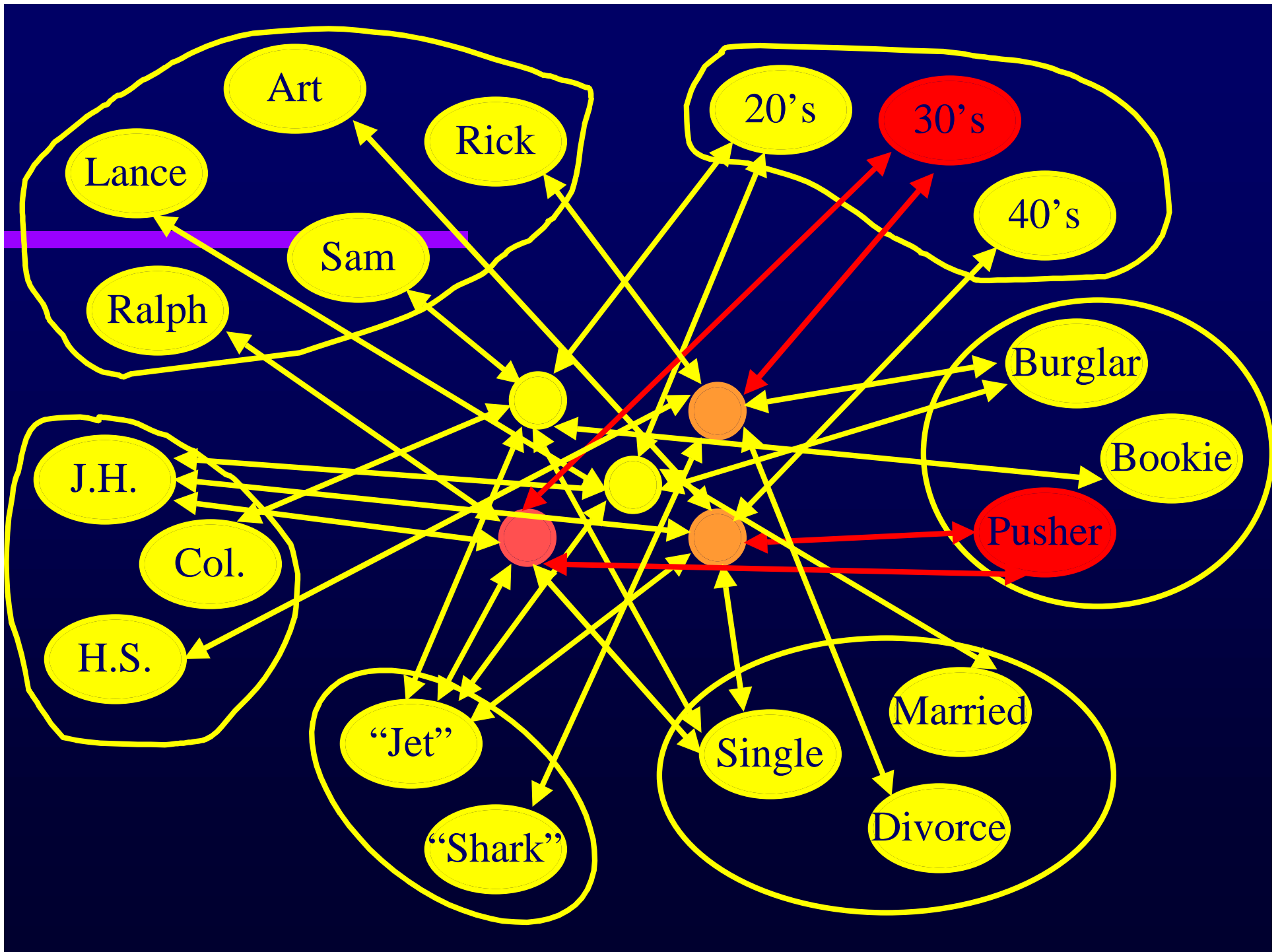


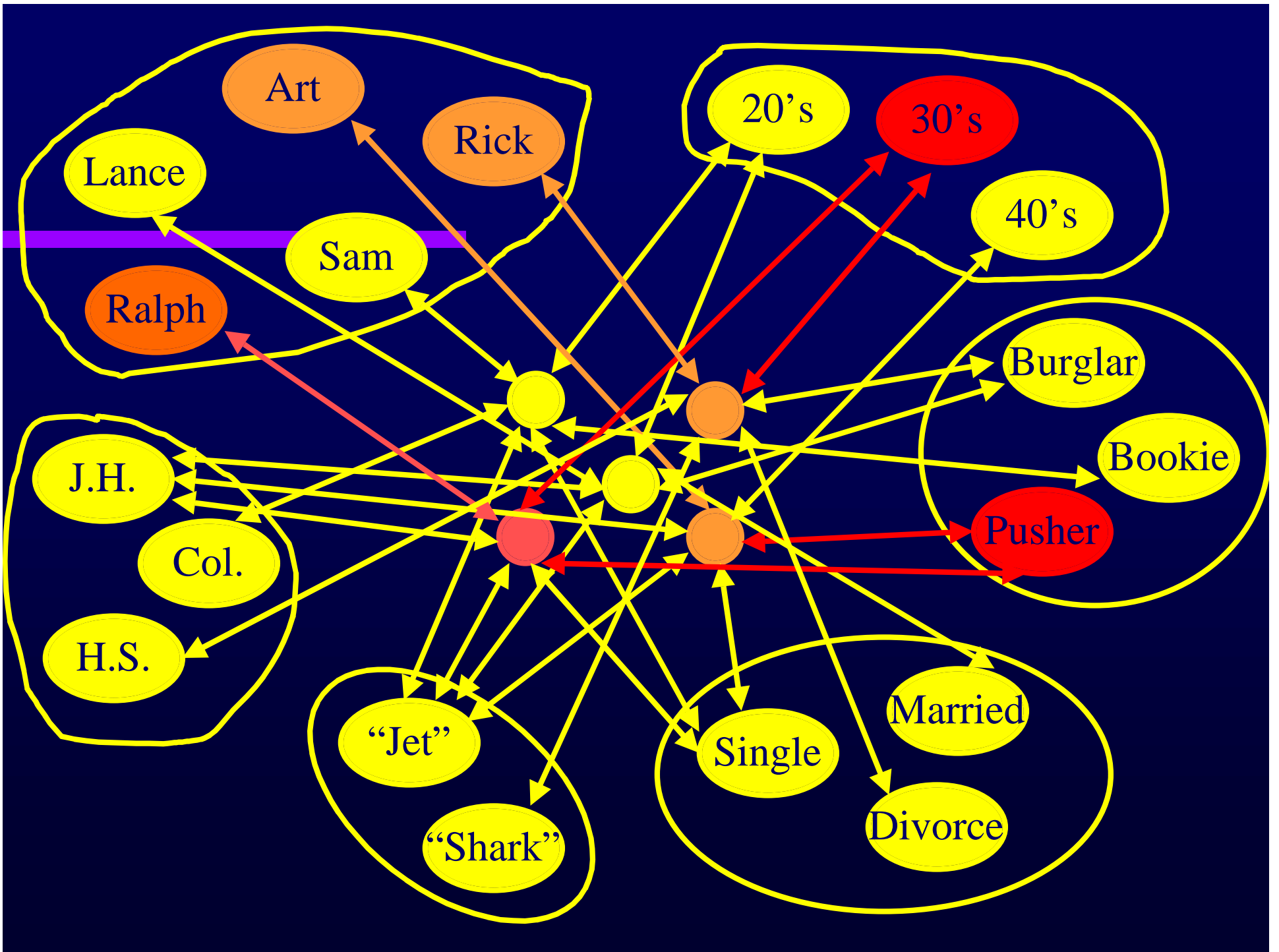


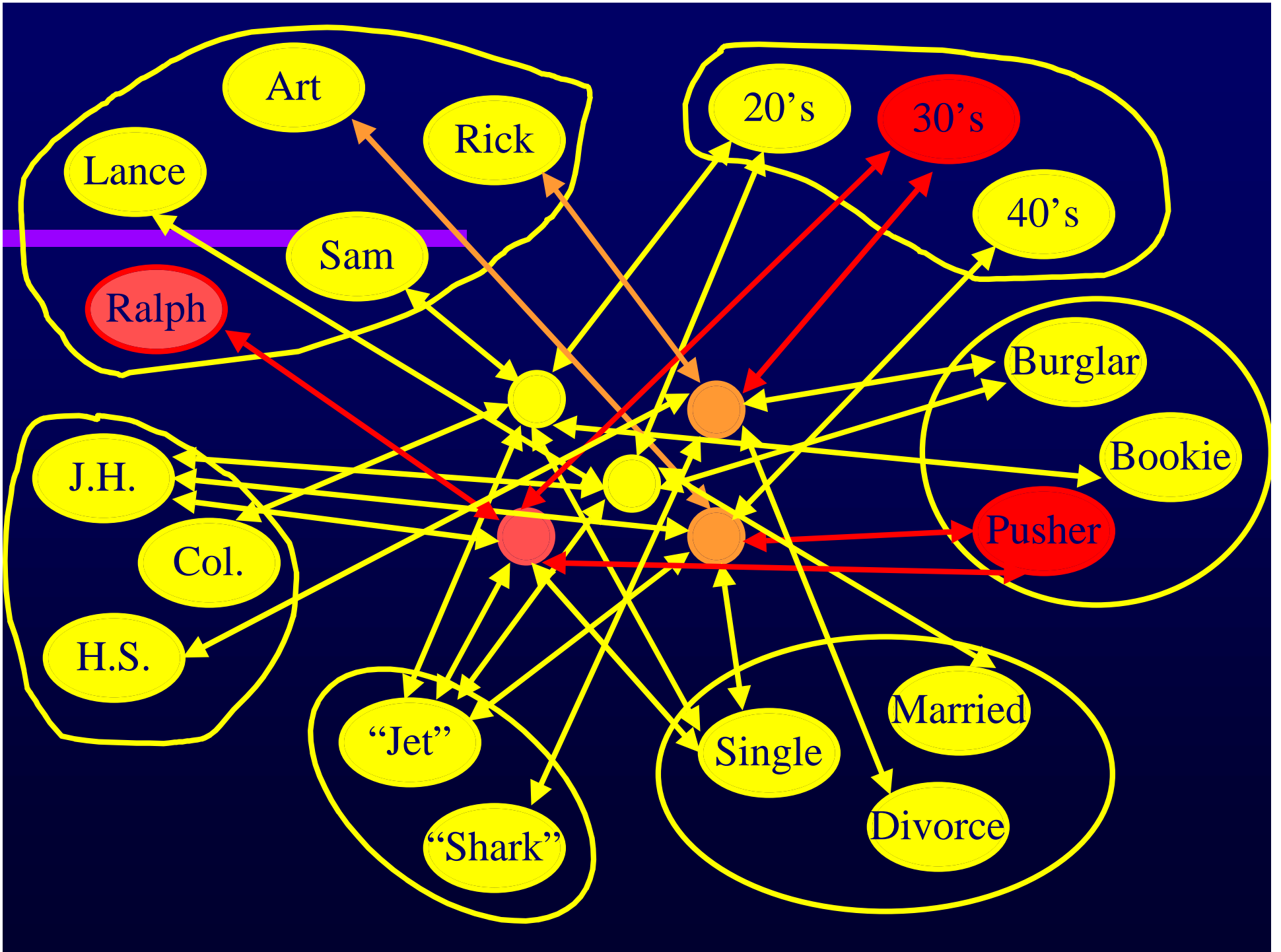
Vraag aan het netwerk

- Wie is in de dertig en is een pusher?



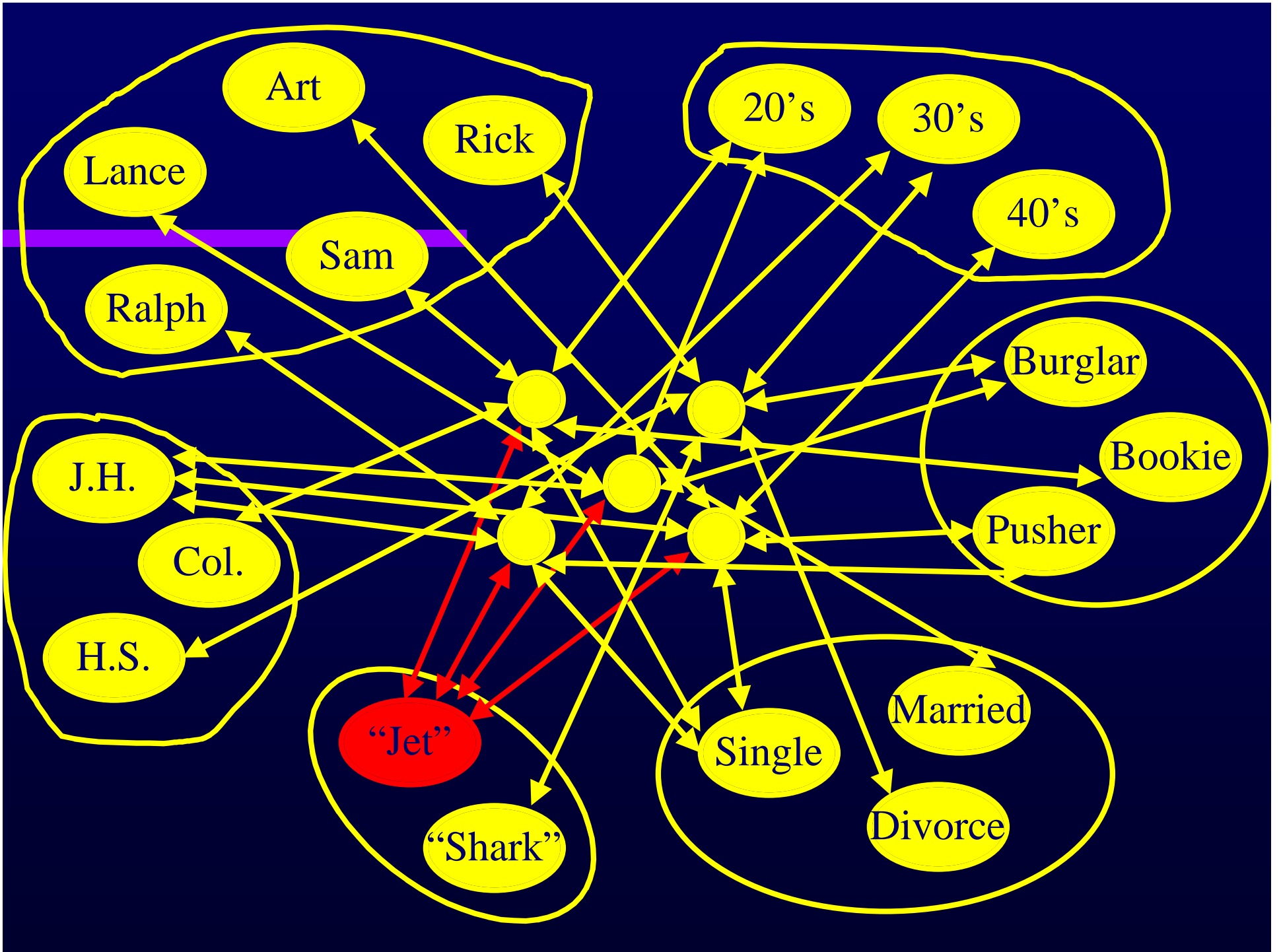


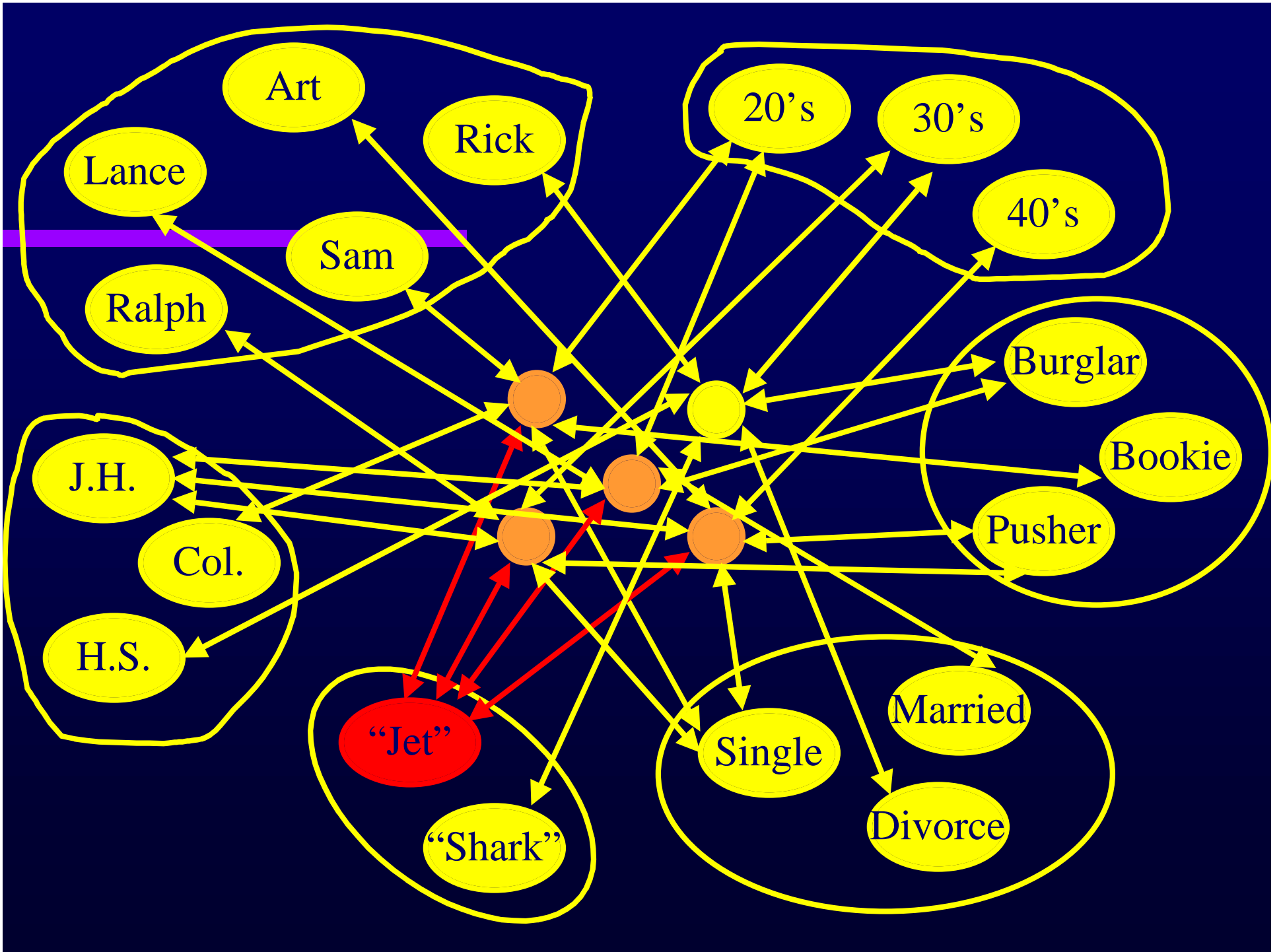


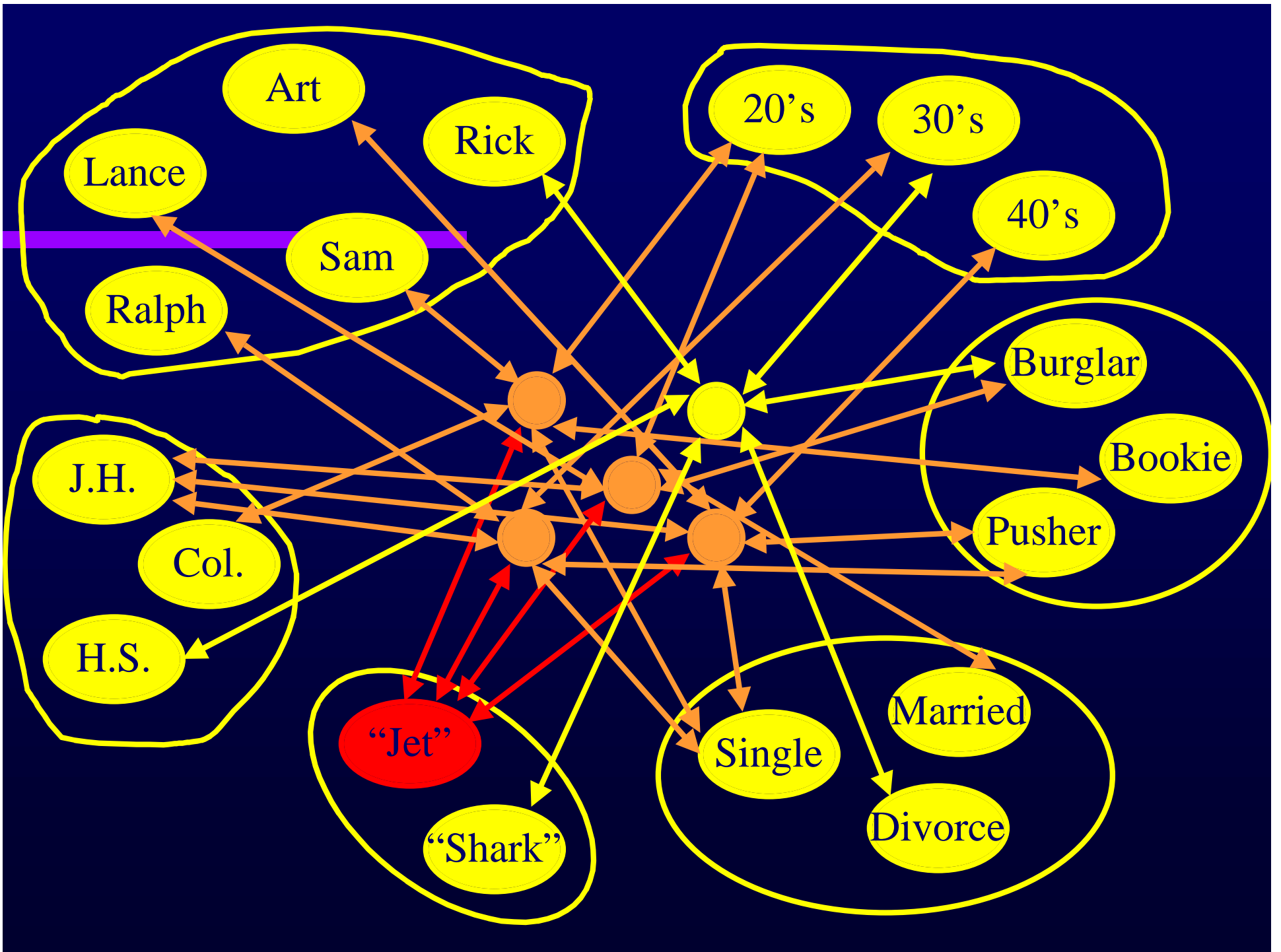


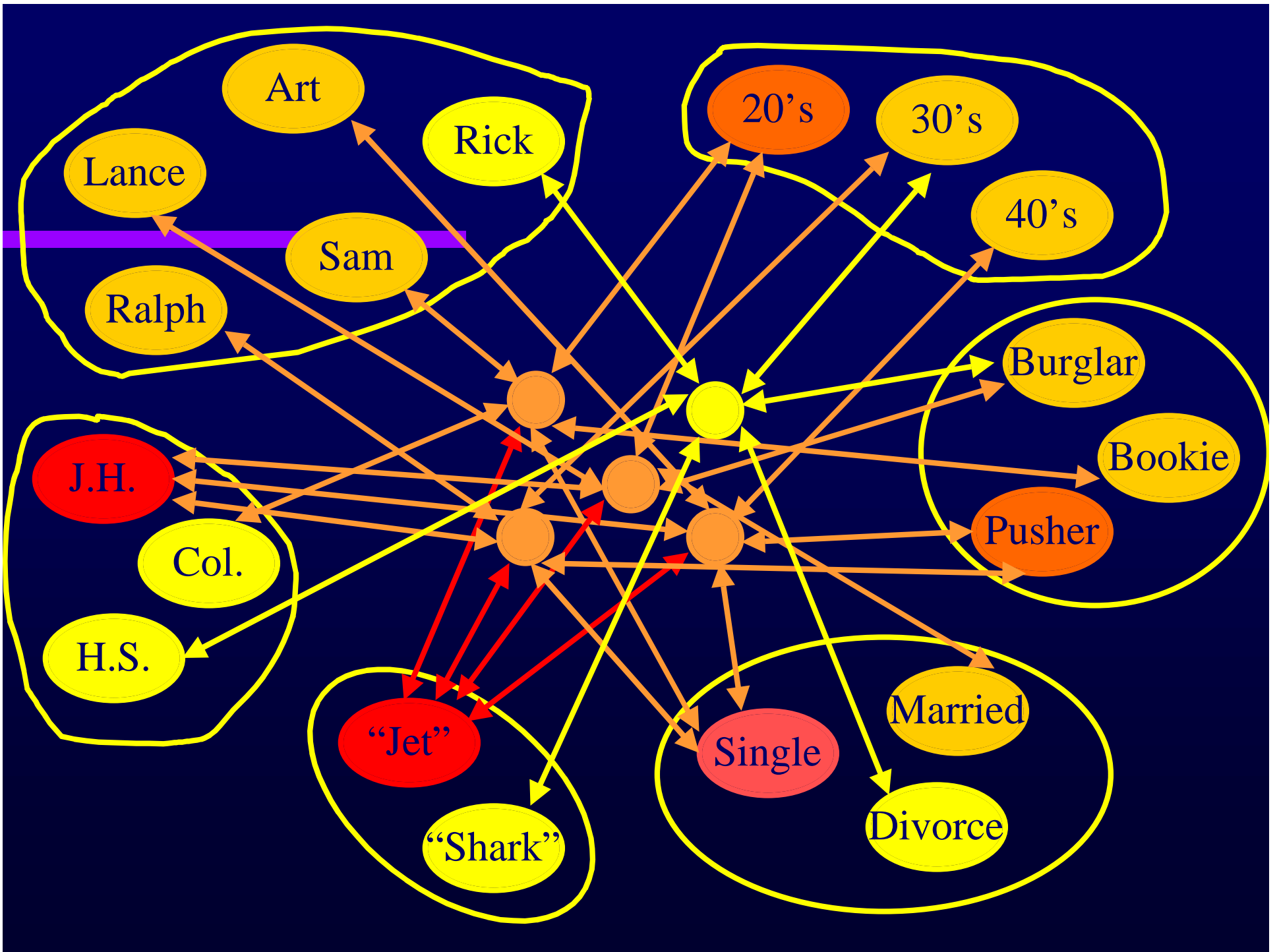
Vraag aan het netwerk

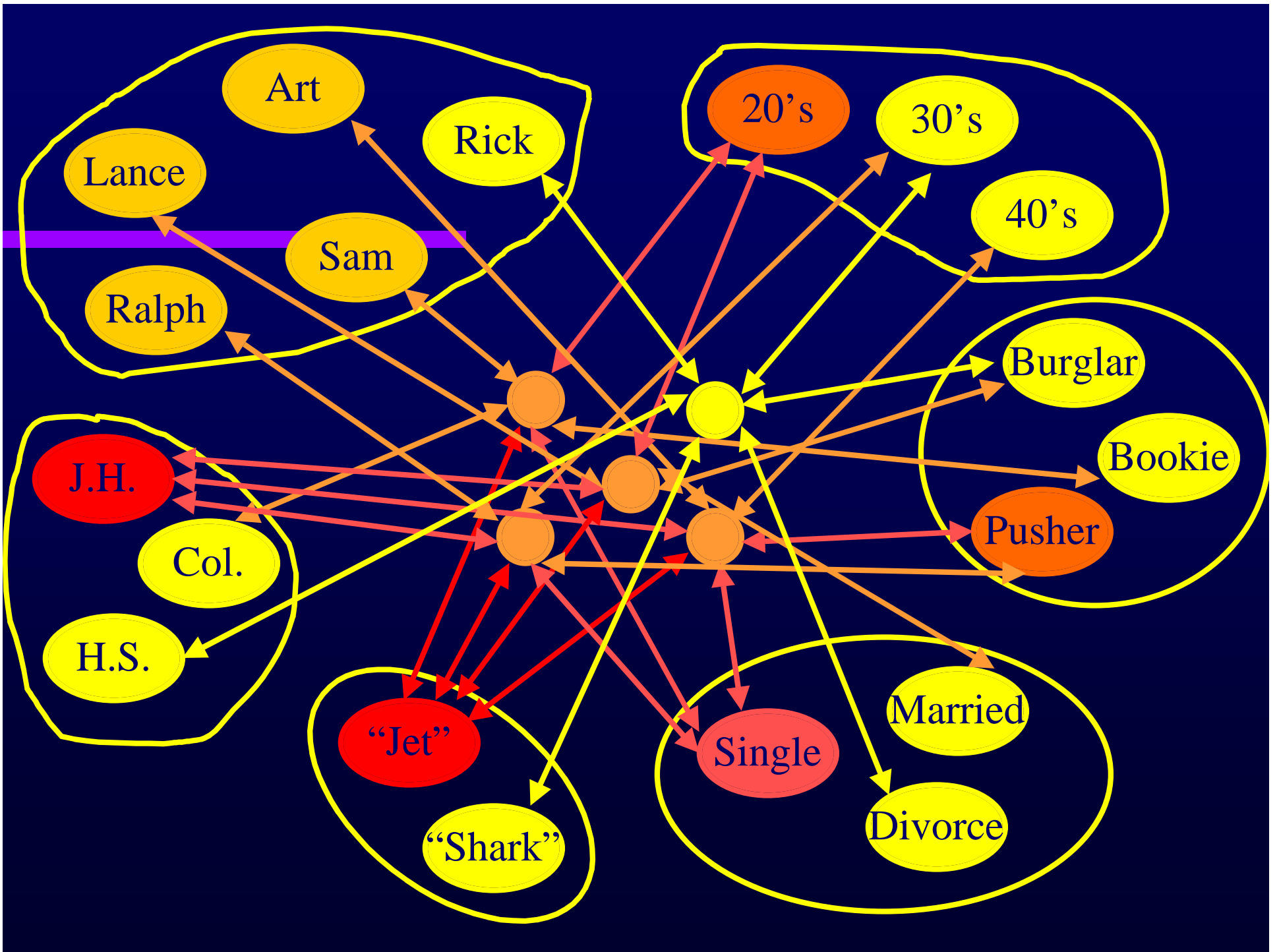
- Wat voor iemand is nu een “Jet”?

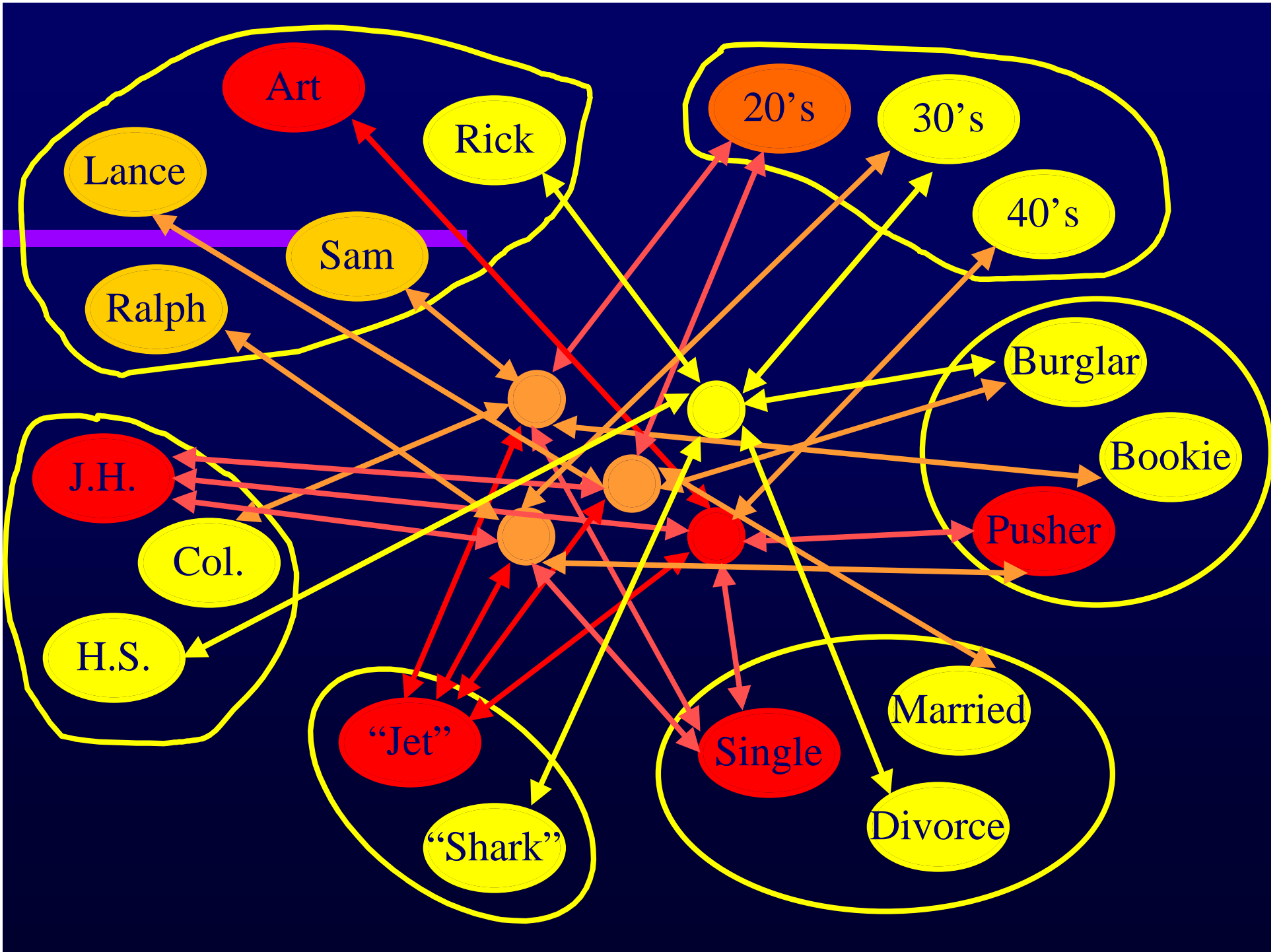






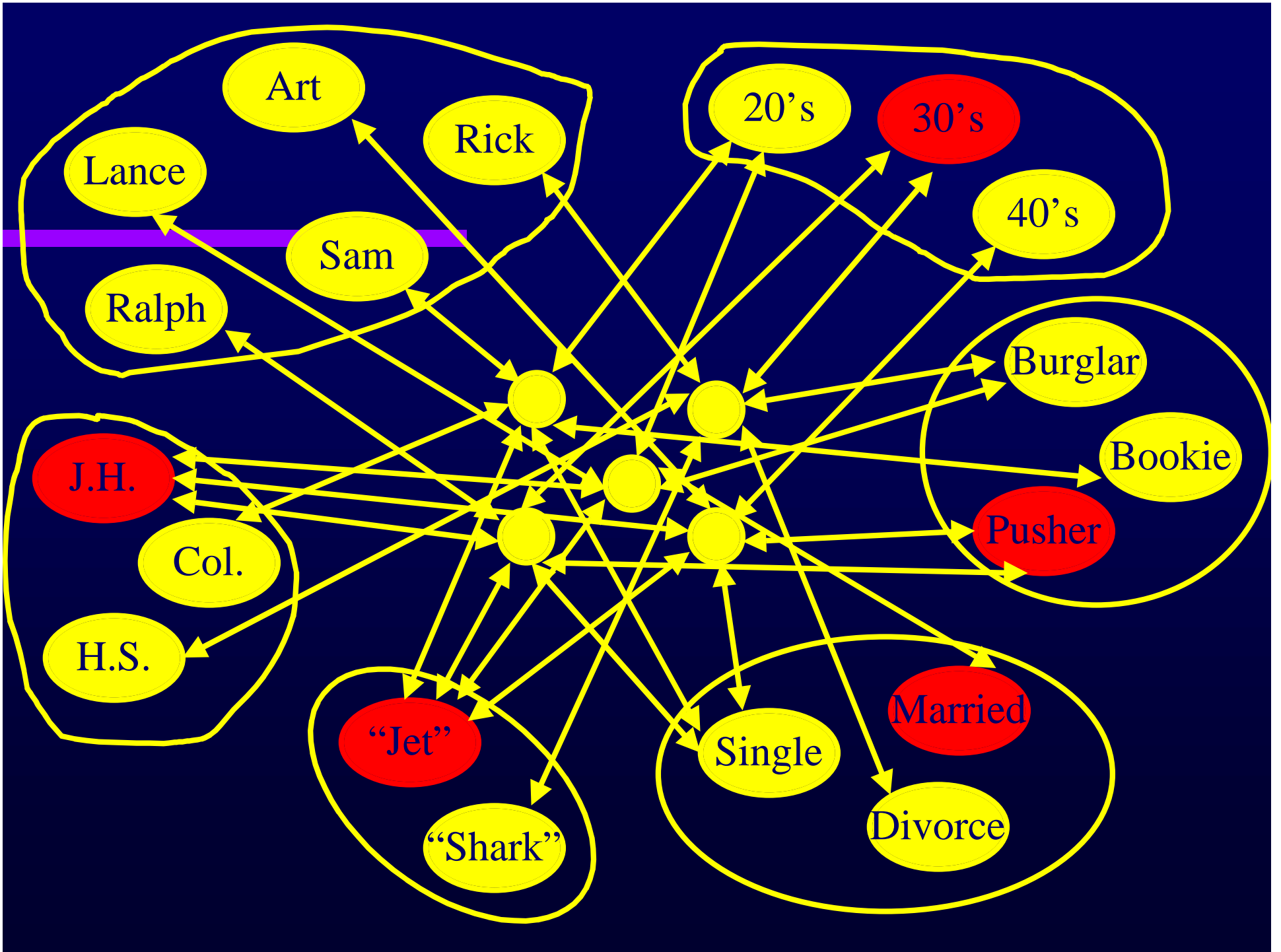


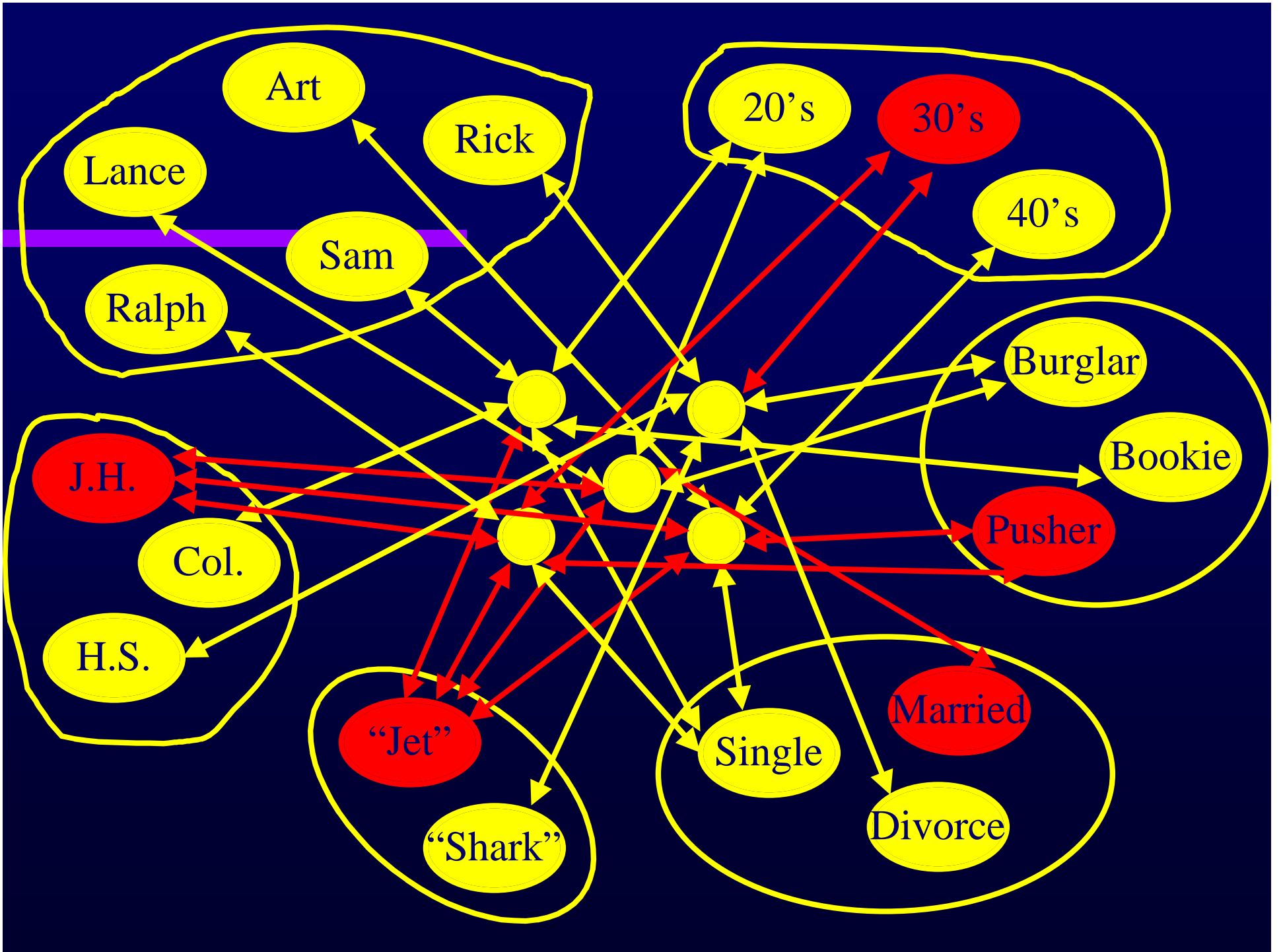


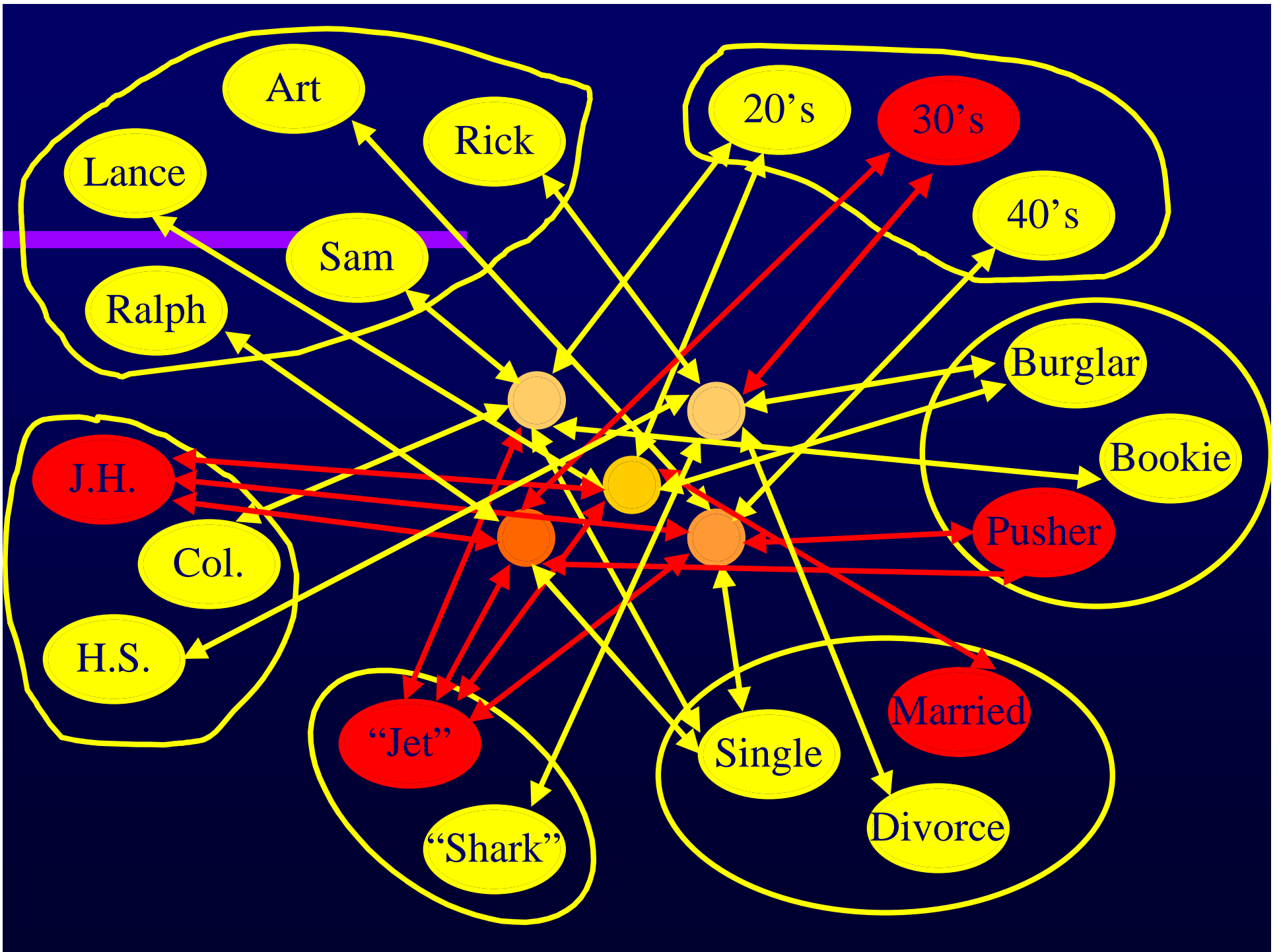


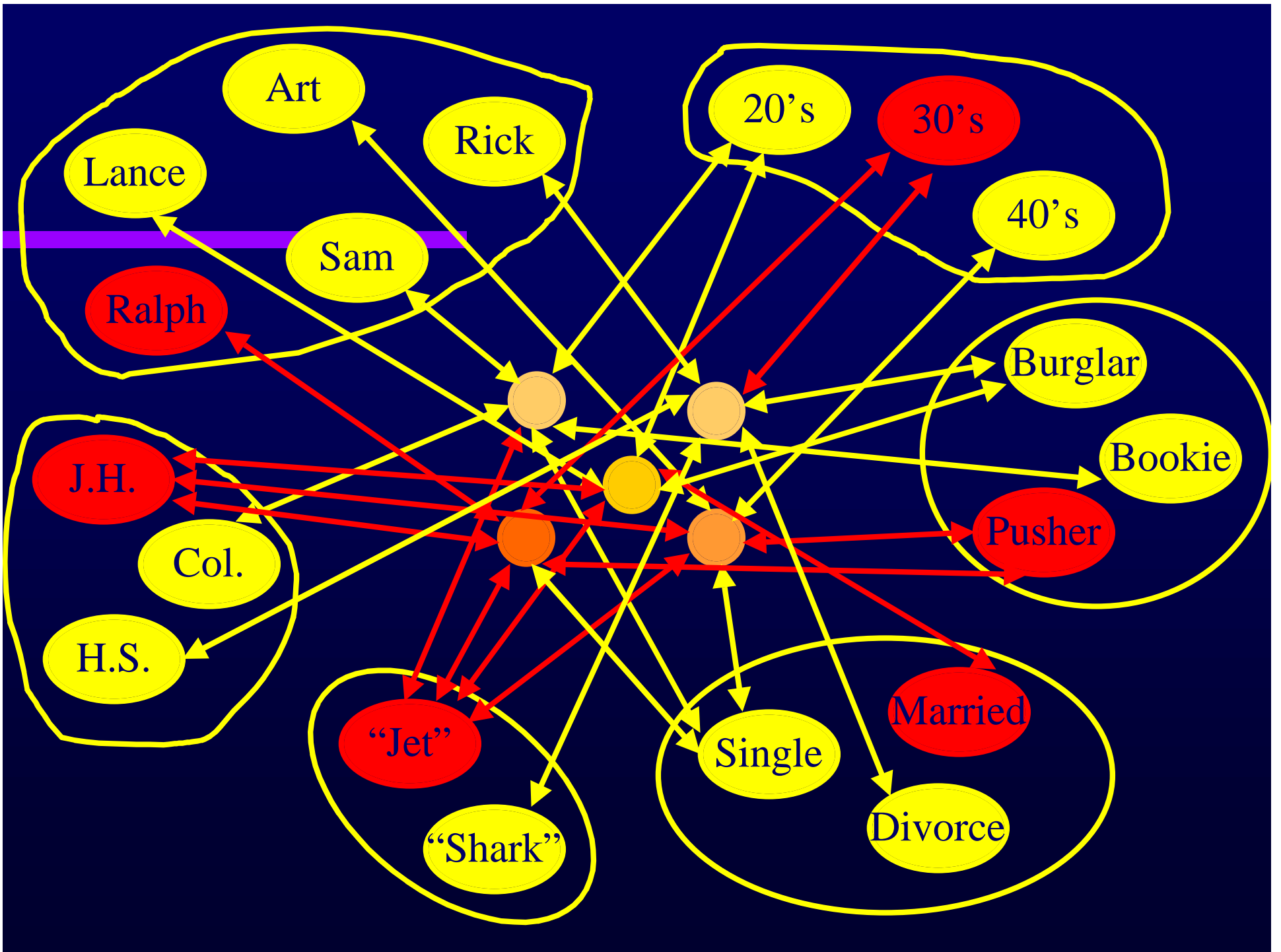
Vraag aan het netwerk

- Wie heeft de volgende eigenschappen:
J.H., Jet, married, pusher en 30's?









Eigenschappen neurale netwerk

- Kan snel dingen opzoeken, door parallelisme
- Kan “prototypes” afleiden: een “typische” Jet
- Kan omgaan met fouten en onvolledigheden in de invoer

Andere eigenschappen (niet in dit voorbeeld)

- Gedistribueerde representatie:
 - Informatie wordt opgeslagen door meer dan één cel en met meer dan één verbinding
 - Neurale netwerken kunnen zelf hun kennis leren door ervaring

ACT-R

- Is een zogenaamde “hybride” architectuur
 - Gedeeltelijk gebruik makend van symbolen
 - Gedeeltelijk gebruik makend van activatiewaarden e.d., hetgeen ook wel “sub-symbolisch” wordt genoemd

ACT-R heeft 2 soorten geheugen

- Een declaratief geheugen: een geheugen voor feiten
- Een procedureel geheugen: een geheugen voor handelingen

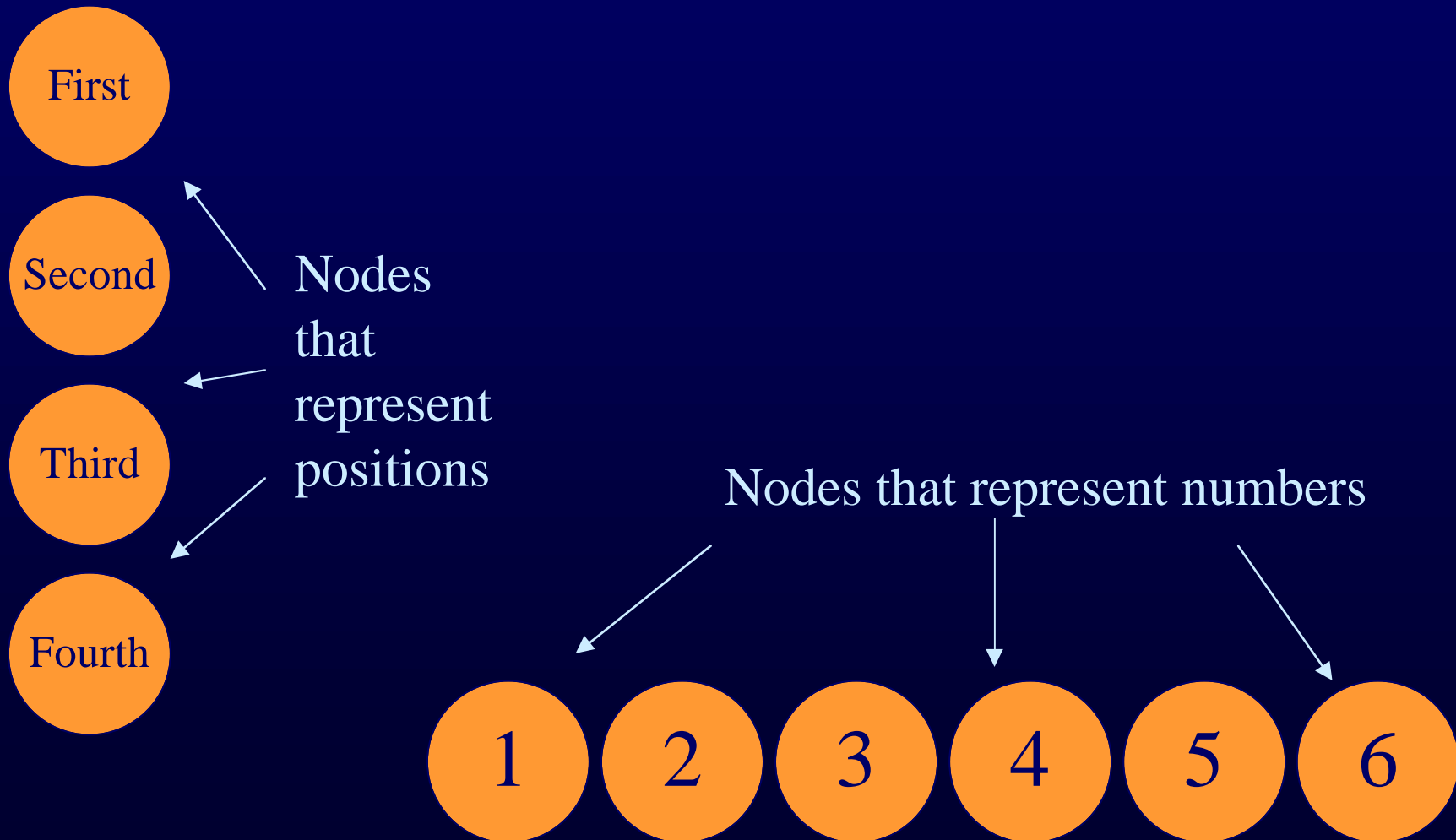
Declaratief geheugen

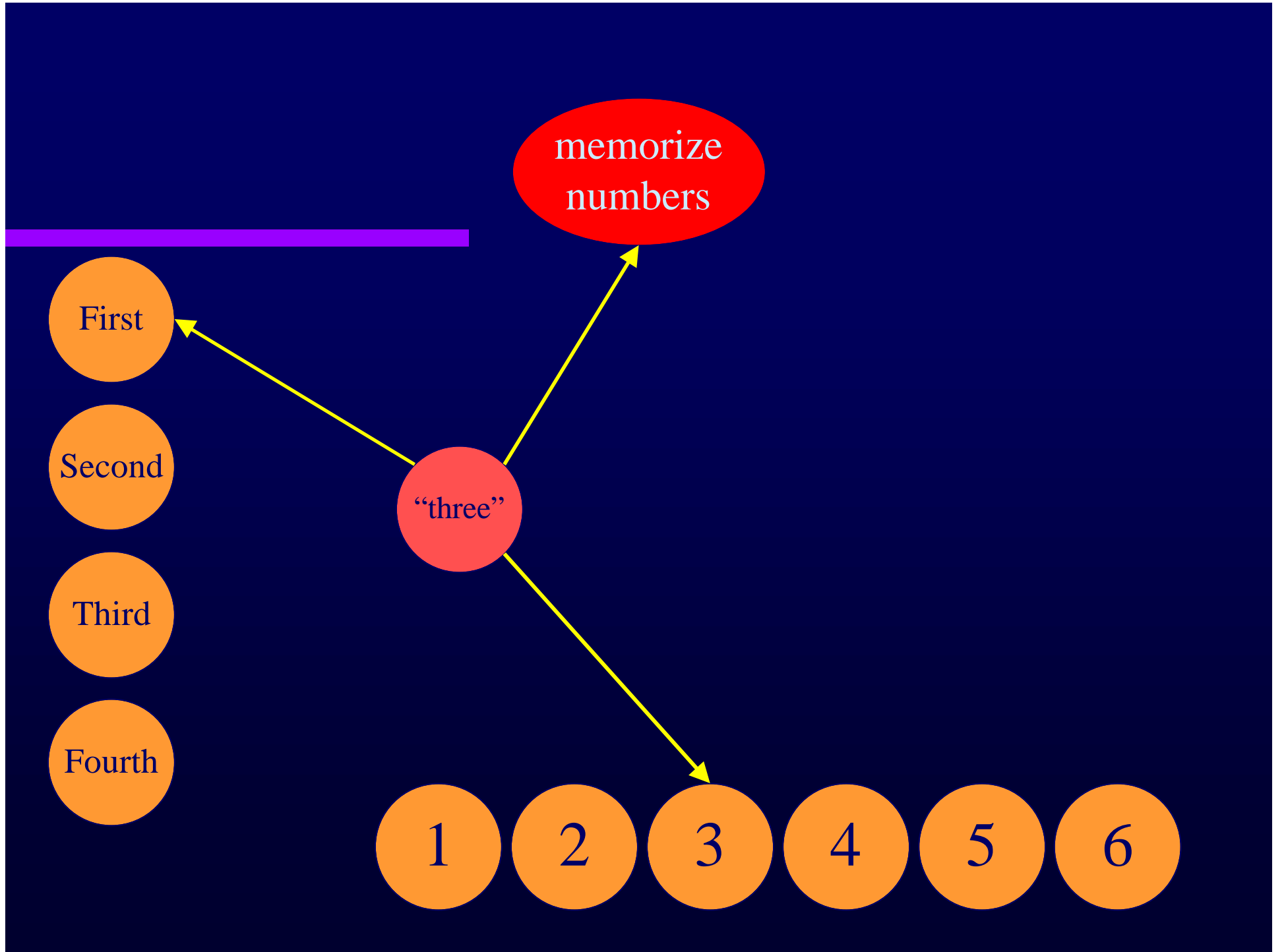
- ACT-R's geheugen voor feiten lijkt wel een beetje op het Jets en Sharks netwerk
 - Voor elk feit is er een knoop (een “chunk”)
 - Chunks zijn met elkaar verbonden
 - Chunks hebben een activatie, die ze kunnen doorgeven via hun verbindingen aan andere chunks

Korte-termijn geheugen

- In veel theorieën over geheugen is sprake van een apart korte-termijn geheugen (ook wel werkgeheugen), bijvoorbeeld de Atkinson en Shiffrin theorie
- Kenmerkend is een **vaste** capaciteit (vaak 7 ± 2)
- Is het **verschijnsel** korte-termijn geheugen echter niet te verklaren zonder het **theoretische construct** korte-termijn geheugen?

Korte-termijn geheugen in het Lange-termijn geheugen





memorize
numbers

First

Second

Third

Fourth

“three”

“six”

1

2

3

4

5

6

memorize
numbers

First

Second

Third

Fourth

“three”

“six”

“two”

1

2

3

4

5

6

memorize
numbers

First

Second

Third

Fourth

“three”

“six”

“two”

“four”

1

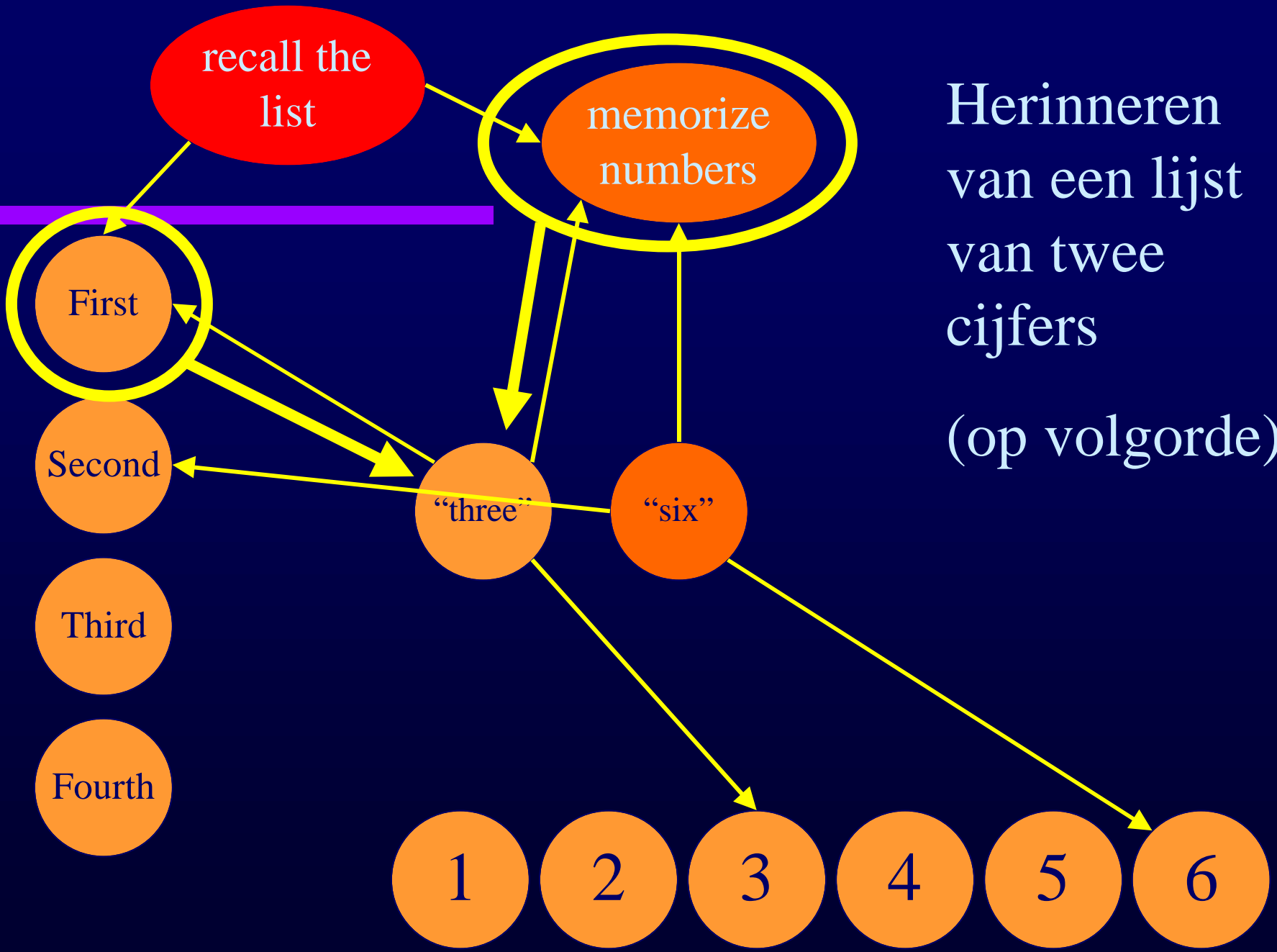
2

3

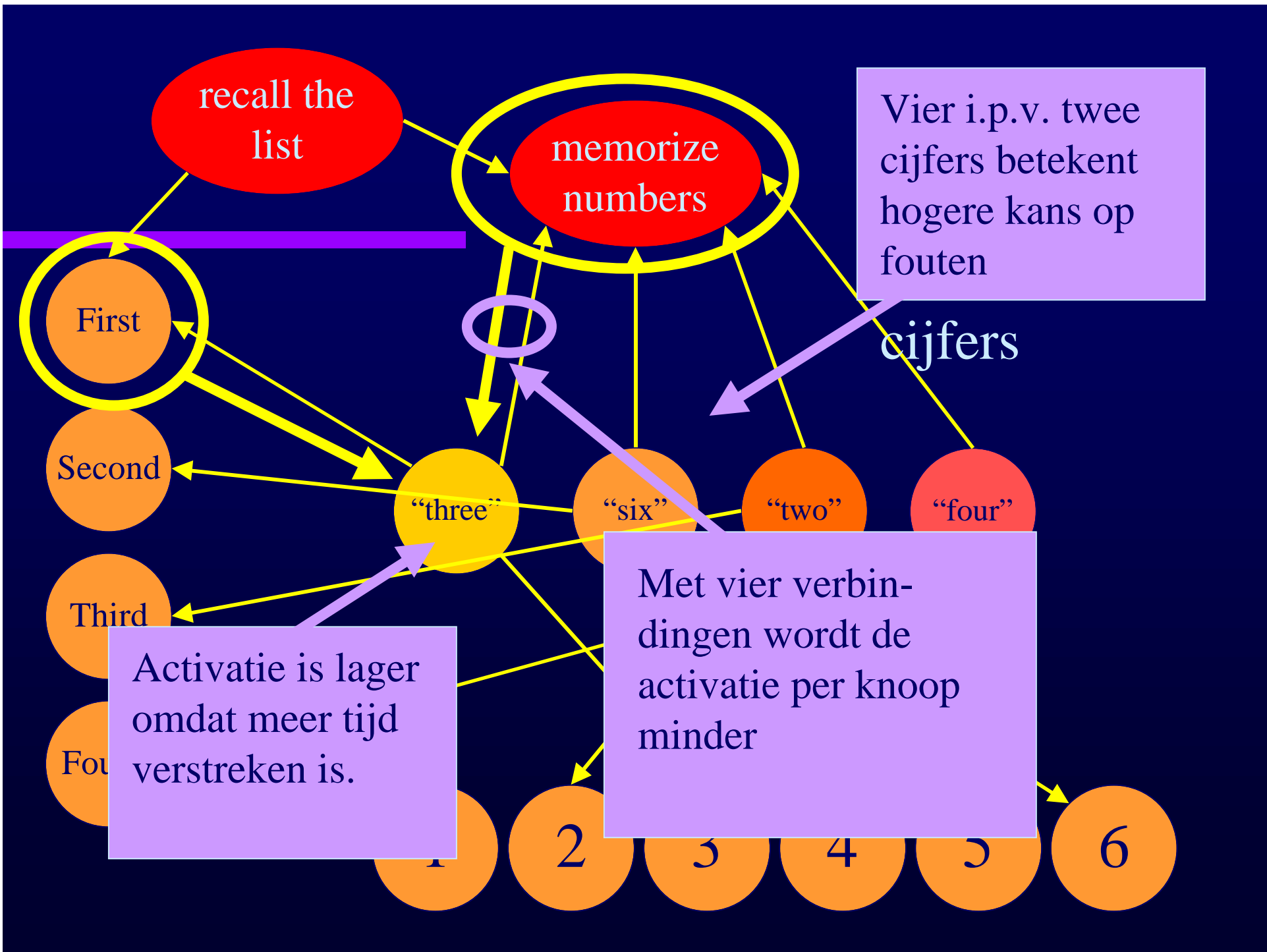
4

5

6

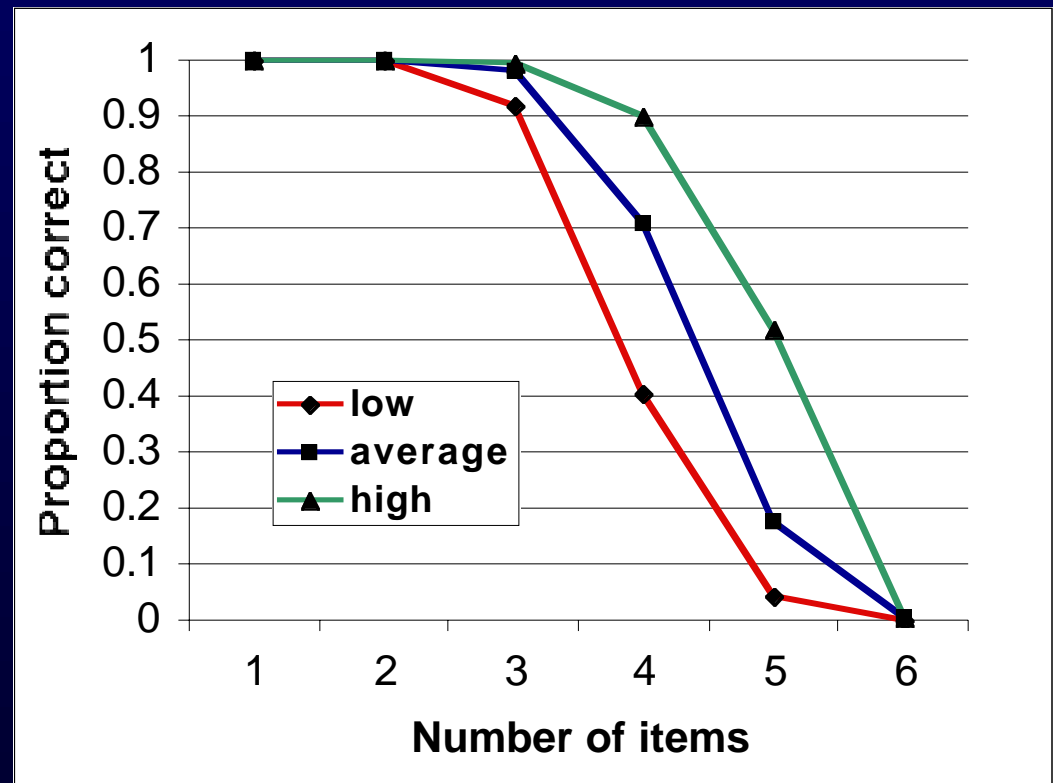


Herinneren
van een lijst
van twee
cijfers
(op volgorde)



Resultaten van dit model

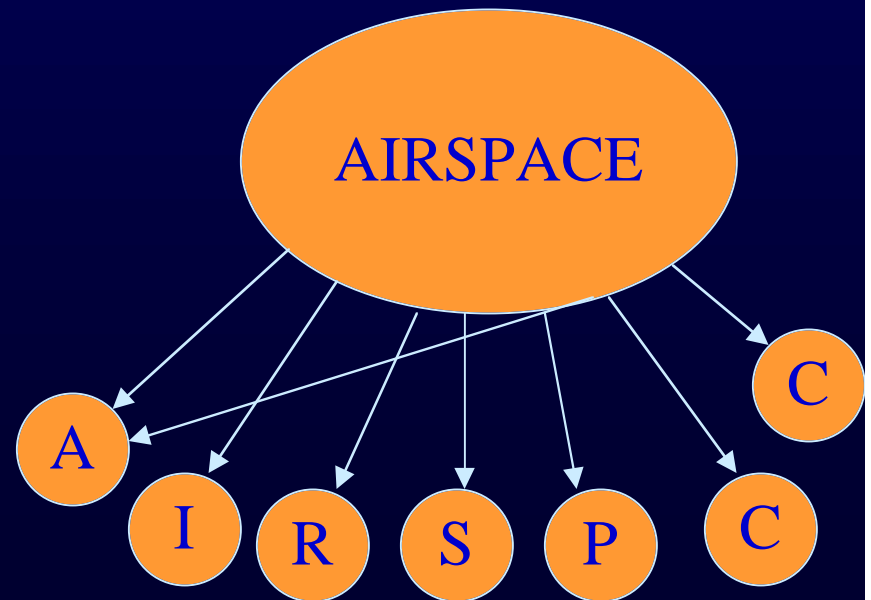
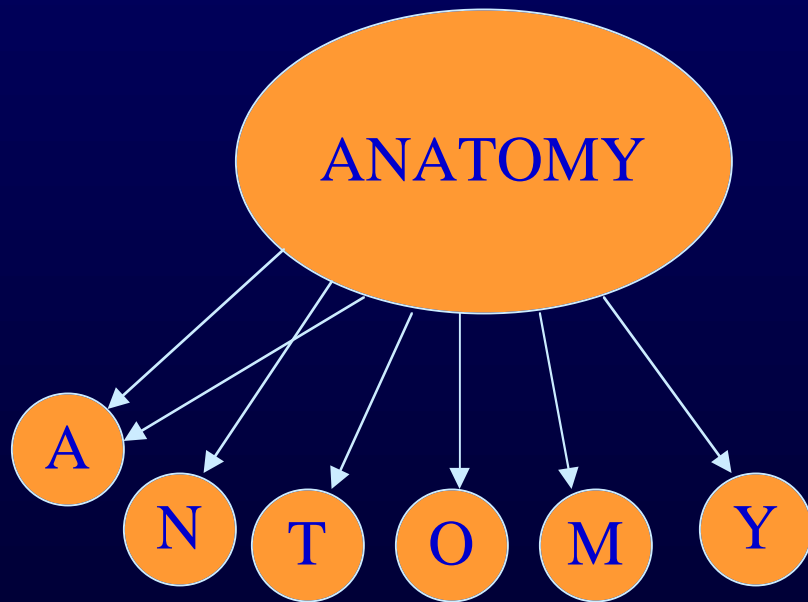
- Volgens dit model is repeteren niet toegestaan
- Als dit wel mag, komt er Miller's zeven \pm twee uit.



Waarom is een theorie zonder Korte-Termijn geheugen beter?

- Hoe meer theoretische constructies, des te zwakker de theorie
- In tegenstelling tot Atkinson & Shiffrin en Squire
 - Is ook echt kennis gerepresenteerd in het geheugen
 - Kunnen geheugenprocessen gesimuleerd worden, zodat we nieuwe inzichten krijgen

Voorbeeld: Tulving's geheugentaak



Experiment

Leerdoel
anatomy

ANATOMY

A
N
T
O
M
Y

AIRSPACE

A
I
R
S
P
C
E

Experiment

Leerdoel
anatomy

ANATOMY

A

N

T

O

M

Y

AIRSPACE

A

I

R

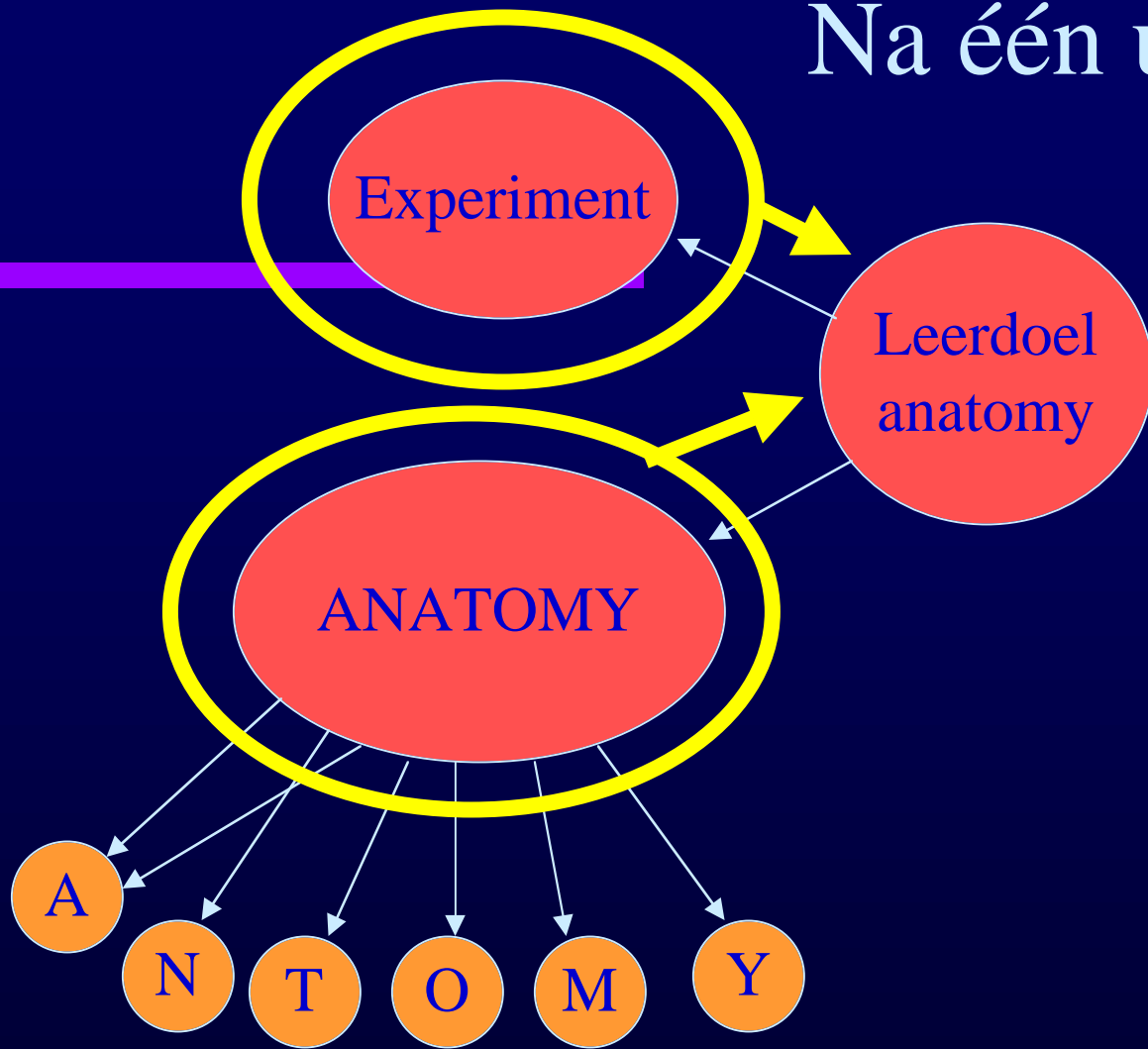
S

P

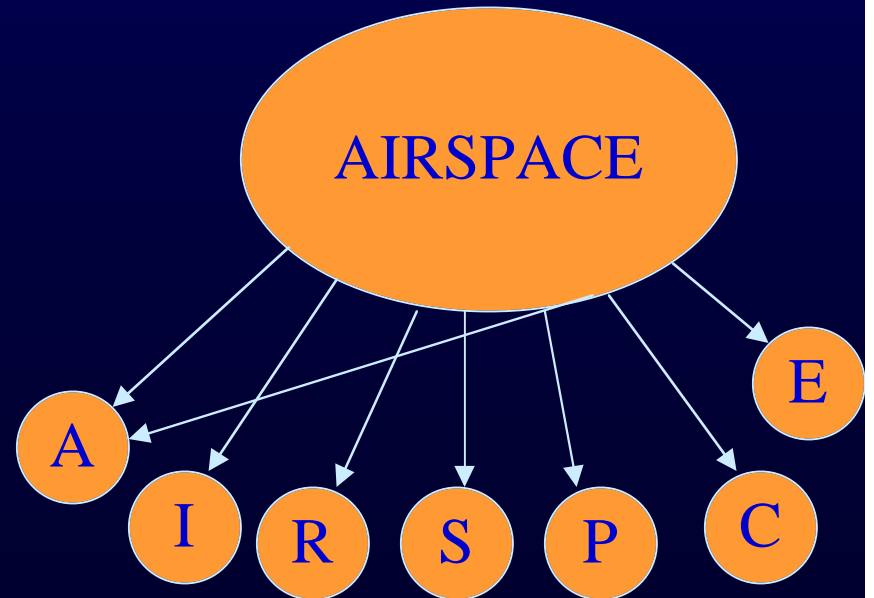
C

E

Na één uur



Stond "Anatomy"
op de lijst?



Na één uur

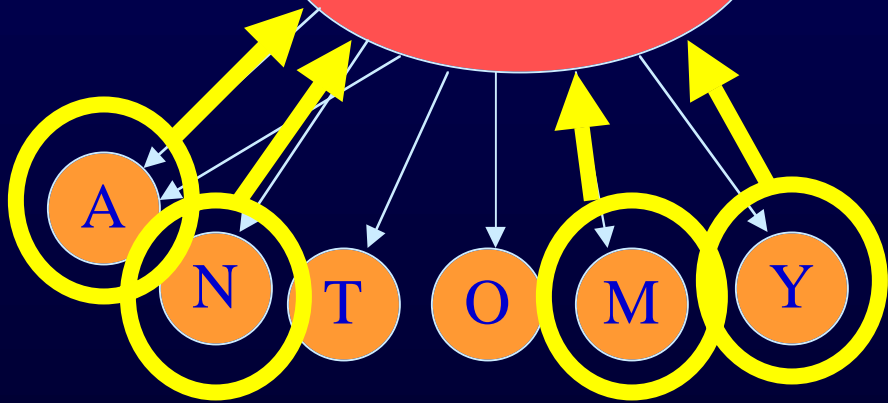
Experiment

Leerdoel
anatomy

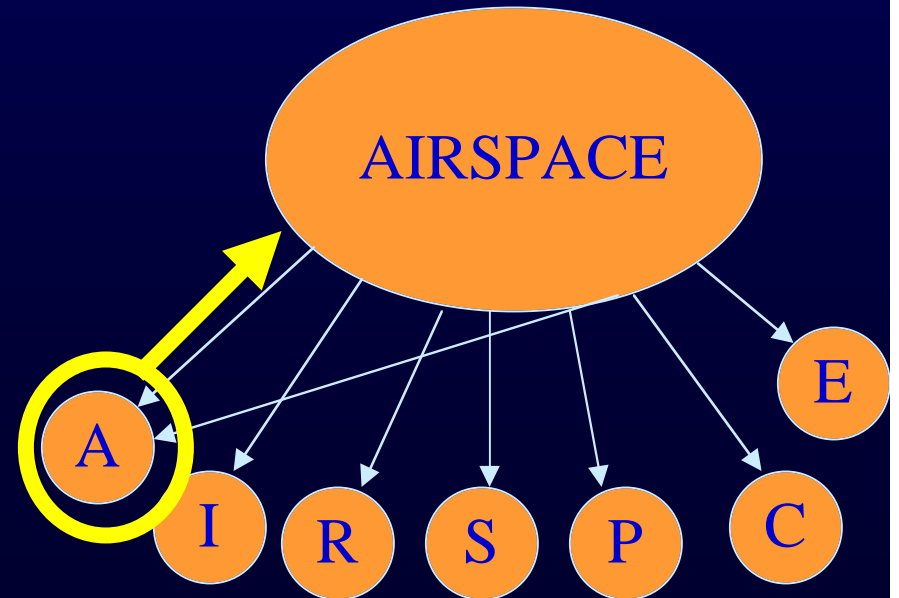
Maak af:

AN___MY

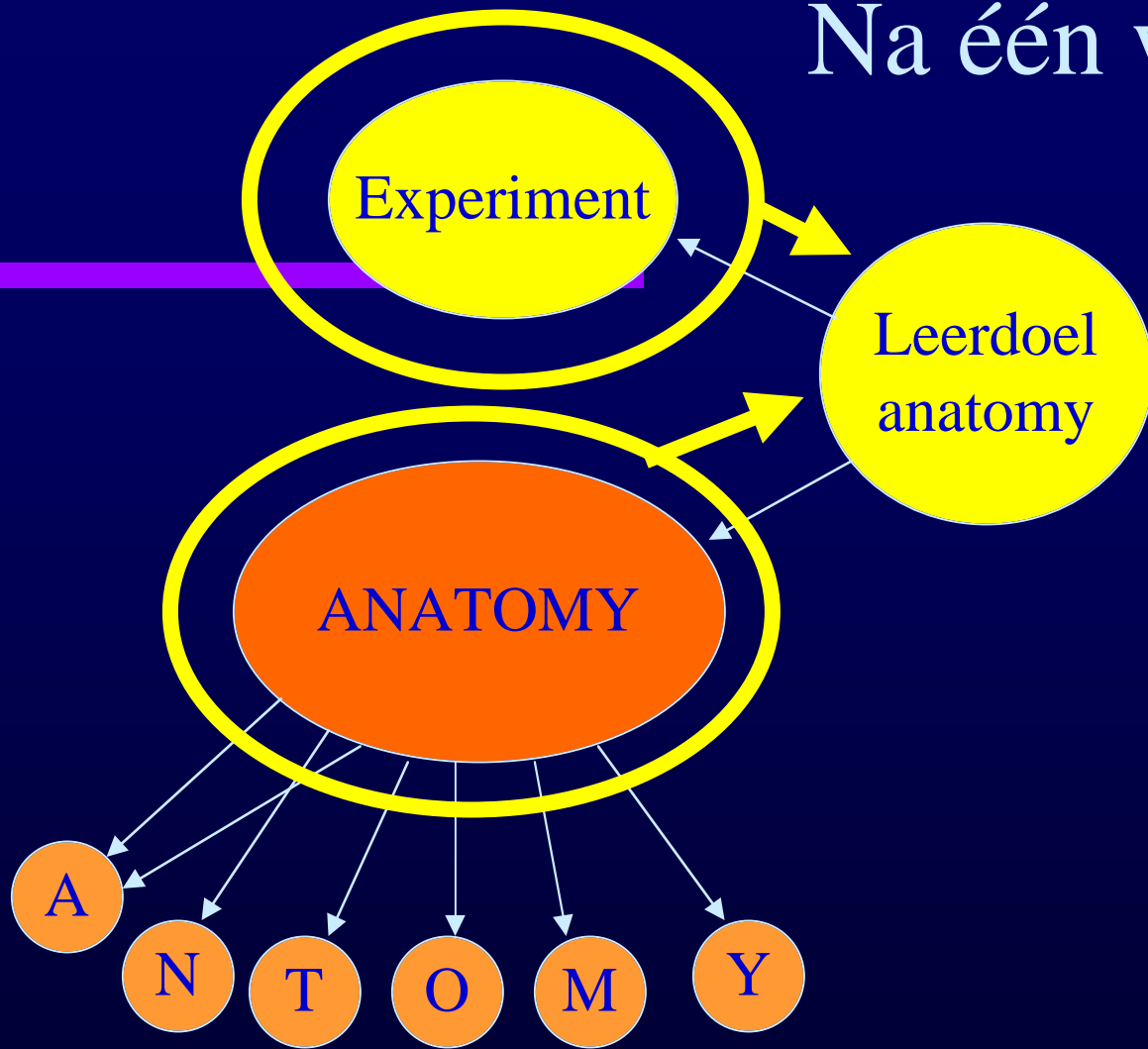
ANATOMY



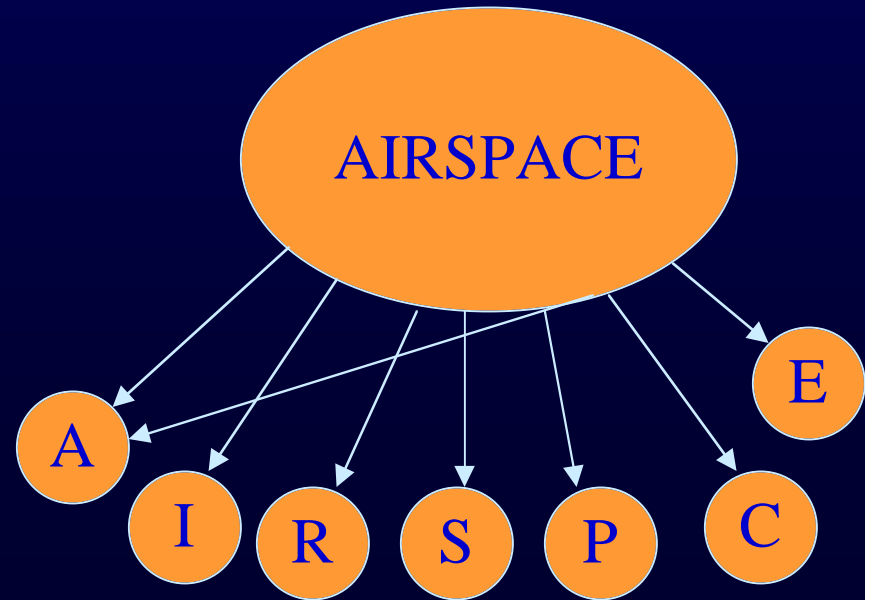
AIRSPACE



Na één week



Stond "Anatomy"
op de lijst?



Na één uur

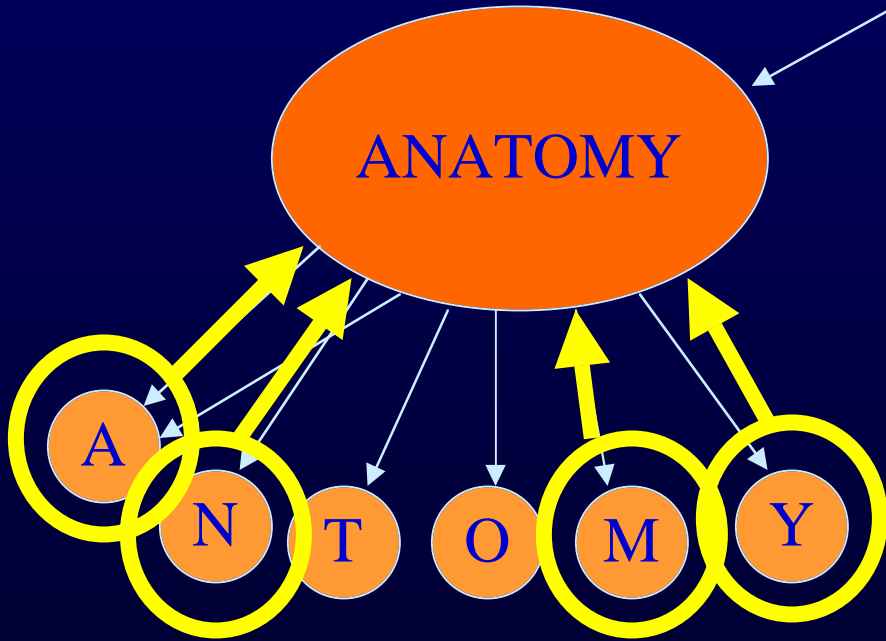
Experiment

Leerdoel
anatomy

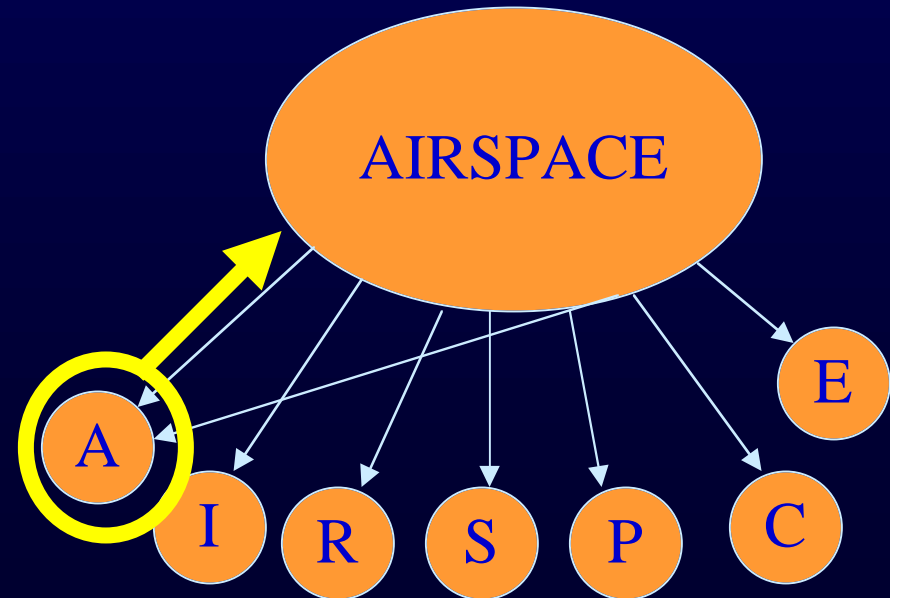
Maak af:

AN___MY

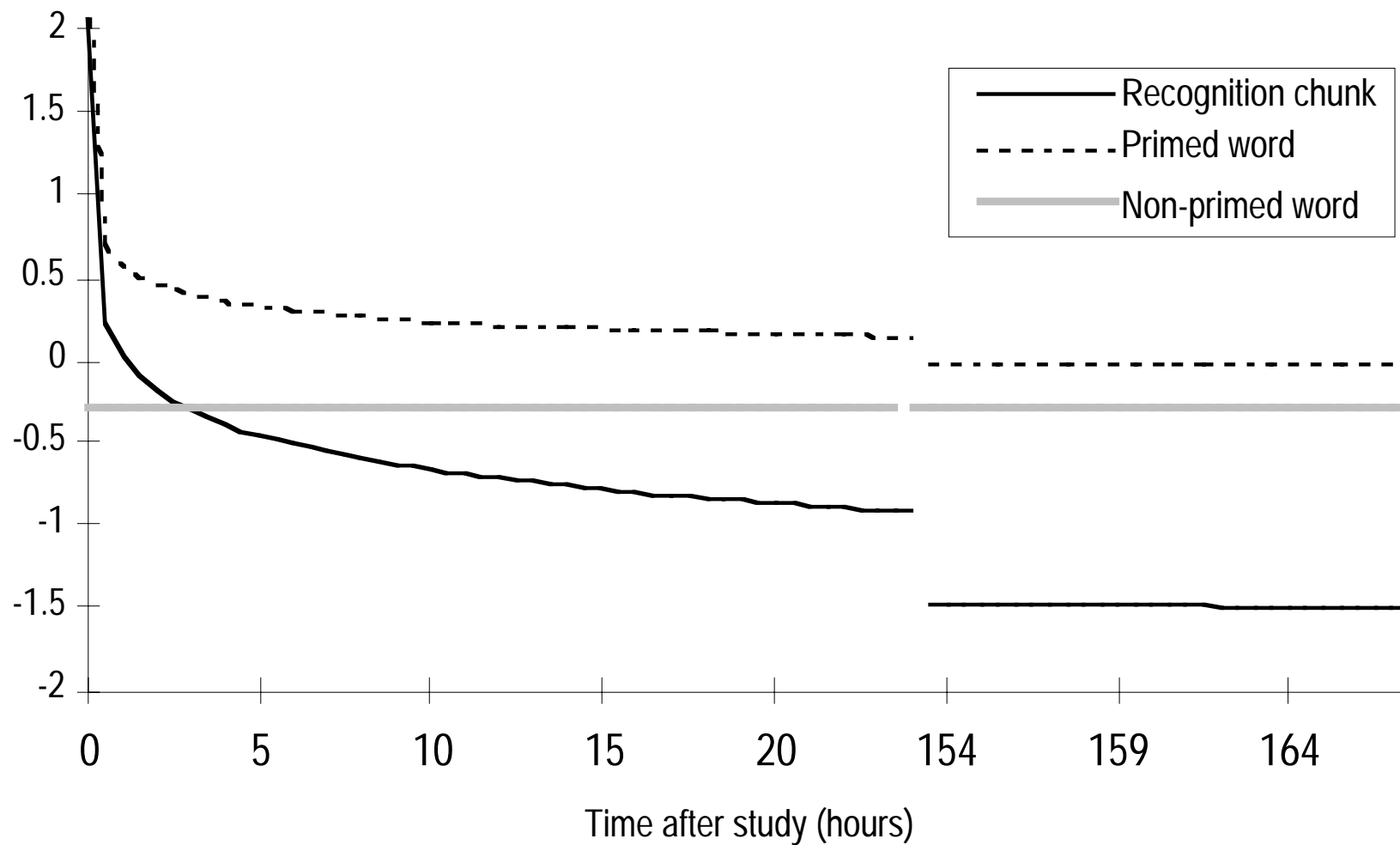
ANATOMY



AIRSPACE

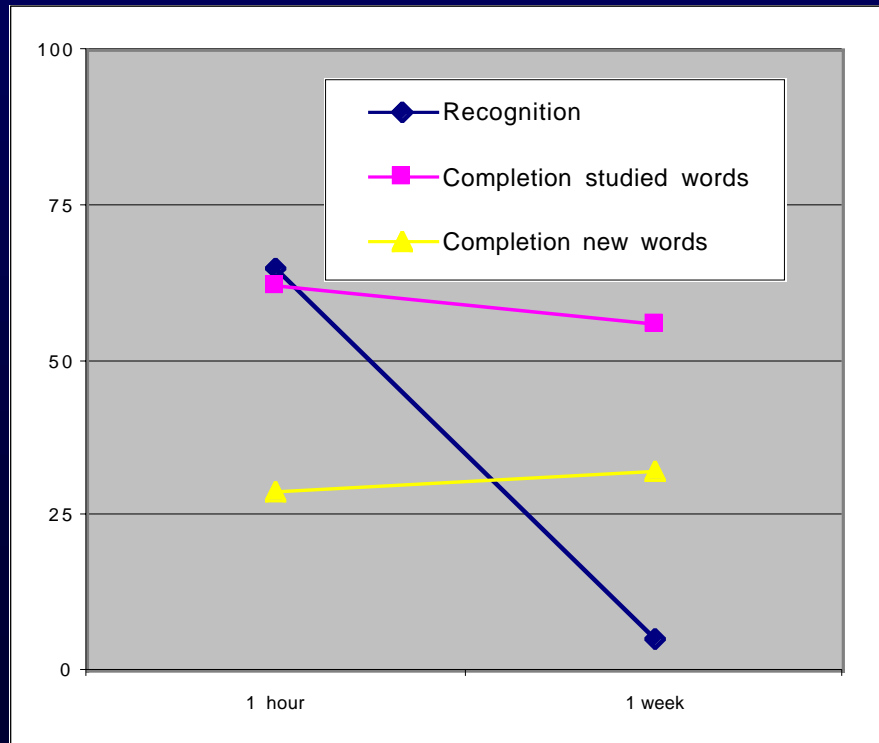


Grafiek van de activatie

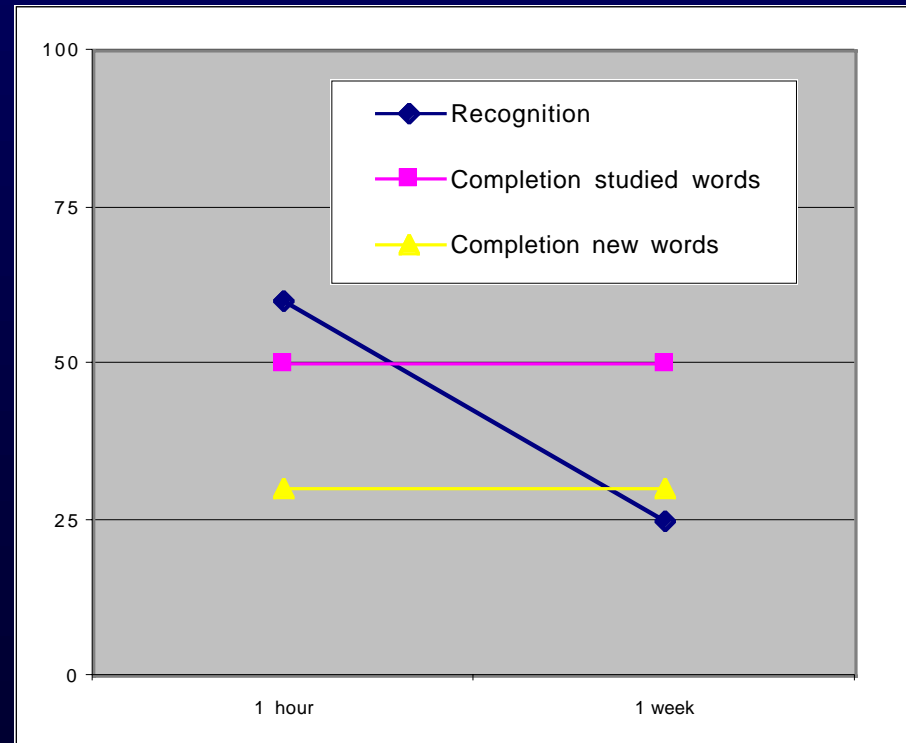


Resultaten van het model

Model



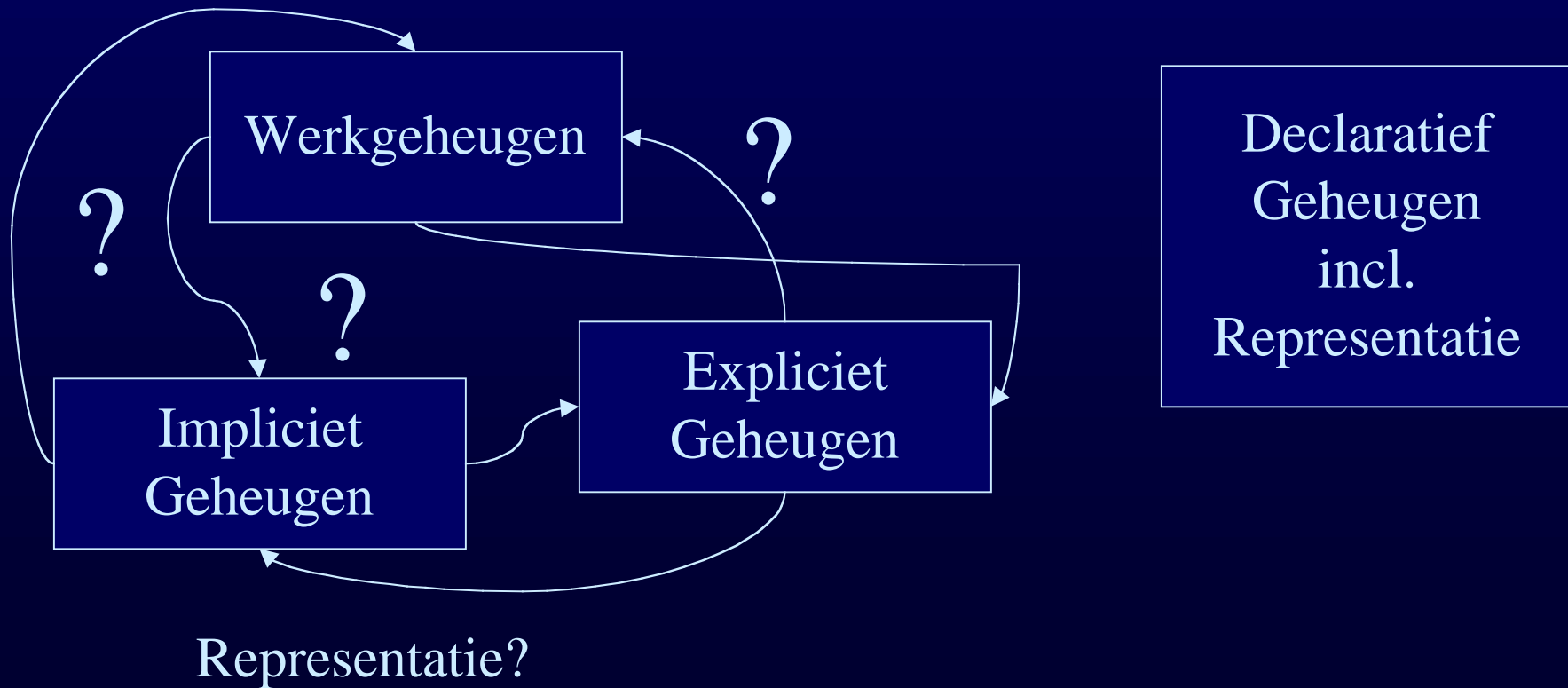
Data



Conclusie: gebruik van formele technieken

Oud:

Nieuw:



Kennis van handelingen in ACT-R

- Door middel van regels
- Voorbeeld in het boek 5.3 (164-167)
 - Aftreksommen in kolommen
 - Negen regels kunnen de hele vaardigheid
 - Zes regels kunnen het zonder lenen
 - Eén regel kan eenvoudige aftreksommen

Waar kan ik deze slides vinden?

- <http://www.ai.rug.nl/~post>