



ugr

Universidad
de **Granada**

TRABAJO FIN DE GRADO
INGENIERÍA EN ...

Análisis de procesos

Subtitulo del Proyecto

Autor

María Isabel Ruiz Martínez (alumno)

Directores

Luis Castillo Vidal (tutor1)

Nombre Apellido1 Apellido2 (tutor2)



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN

Granada, mes de 201



Análisis de procesos

Subtítulo del proyecto.

Autor

María Isabel Ruiz Martínez (alumno)

Directores

Luis Castillo Vidal (tutor1)

Nombre Apellido1 Apellido2 (tutor2)

Análisis de procesos: Subtítulo del proyecto

Nombre Apellido1 Apellido2 (alumno)

Palabras clave: palabra_clave1, palabra_clave2, palabra_clave3,

Resumen

Poner aquí el resumen.

Project Title: Project Subtitle

First name, Family name (student)

Keywords: Keyword1, Keyword2, Keyword3,

Abstract

Write here the abstract in English.

Yo, **María Isabel Ruiz Martínez**, alumno de la titulación **TITULACIÓN de la Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación de la Universidad de Granada**, con DNI 75576979Z, autorizo la ubicación de la siguiente copia de mi Trabajo Fin de Grado en la biblioteca del centro para que pueda ser consultada por las personas que lo deseen.

Fdo: María Isabel Ruiz Martínez

Granada a X de mes de 201 .

D. **Luis Castillo Vidal (tutor1)**, Profesor del Área de XXXX del Departamento YYYY de la Universidad de Granada.

D. **Nombre Apellido1 Apellido2 (tutor2)**, Profesor del Área de XXXX del Departamento YYYY de la Universidad de Granada.

Informan:

Que el presente trabajo, titulado ***Análisis de procesos, Subtítulo del proyecto***, ha sido realizado bajo su supervisión por **María Isabel Ruiz Martínez (alumno)**, y autorizamos la defensa de dicho trabajo ante el tribunal que corresponda.

Y para que conste, expiden y firman el presente informe en Granada a X de mes de 201 .

Los directores:

Luis Castillo Vidal (tutor1) **Nombre Apellido1 Apellido2 (tutor2)**

Agradecimientos

Poner aquí agradecimientos...

Índice general

1. Introducción, motivaciones, objetivos y estructura	1
1.1. Introducción	1
1.2. Motivaciones	2
1.3. Objetivos	2
1.4. Organización de este Trabajo Fin de Grado	2
2. Planificación	5
2.1. Herramientas	5
3. Análisis	7
3.1. Algunos conceptos relevantes	7
3.2. Planteamiento del problema	9
4. Diseño	11
5. Implementación	13
5.1. Análisis ANOVA y primeras conclusiones	13
5.1.1. Descripción del dataset	13
5.1.2. Introducción	16
5.1.3. Two way ANOVA	17
5.1.4. Segmentación de los datos	19
5.2. Extracción de procesos ocultos en los datasets	21
5.2.1. Extracción de los procesos con DISCO	21
5.2.2. Segmentación por años	26
5.2.3. Segmentación por calificaciones	30
5.2.4. Segmentación por año y calificación	36
5.3. Análisis de los procesos con segmentación	54
5.3.1. Arquitectura del software	54
5.3.2. Ejecución del software	54
5.4. Análisis de los motivos en los cambios de comportamiento	54
5.5. Análisis de si hay cambios inducidos por los éxitos de los demás grupos	54

Índice de figuras

3.1. Ejemplo de grafo con 5 nodos.	7
3.2. Ejemplo de grafo dirigido con 5 nodos.	8
3.3. Ejemplo de grafo acíclico dirigido con 5 nodos.	8
5.1. Diagramas de caja y bigotes de las variables <i>map</i> y <i>action</i> . .	14
5.2. Distribuciones para diferentes niveles de los factores	14
5.3. Distribuciones de las variables <i>map</i> y <i>action</i> dependiendo del valor de <i>year</i>	15
5.4. Interacción entre las variables <i>map</i> y <i>action</i>	18
5.5. Gráficas diagnósticas del modelo ANOVA	18
5.6. Análisis de procesos del dataset compuesto (acción mapa). . .	22
5.7. Análisis de procesos del dataset compuesto (acción mapa). . .	23
5.8. Análisis de procesos del dataset compuesto (acción compuesta). .	24
5.9. Análisis de procesos del dataset compuesto (acción compuesta). .	25
5.10. Extracción de procesos de los diferentes cursos académicos (acción mapa/problema).	26
5.11. Extracción de procesos del curso académico 1516 (acción com- puesta).	27
5.12. Extracción de procesos del curso académico 1617 (acción com- puesta).	28
5.13. Extracción de procesos del curso académico 1718 (acción com- puesta).	29
5.14. Extracción de procesos del curso académico 1920 (acción com- puesta).	29
5.15. Extracción de procesos del dataset integrado por las acciones mapa de los grupos según sus calificaciones.	31
5.16. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>baja</i> ”.	32
5.17. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>media-baja</i> ”. . . .	33
5.18. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>media-alta</i> ”. . . .	34

5.19. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>alta</i> ”.	35
5.20. Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “ <i>baja</i> ” a lo largo de cuatro cursos académicos.	36
5.21. Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “ <i>media-baja</i> ” a lo largo de cuatro cursos académicos.	37
5.22. Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “ <i>media-alta</i> ” a lo largo de cuatro cursos académicos.	39
5.23. Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “ <i>alta</i> ” a lo largo de cuatro cursos académicos.	40
5.24. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>baja</i> ” del curso académico 1617.	41
5.25. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>baja</i> ” del curso académico 1718.	42
5.26. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>baja</i> ” del curso académico 1920.	43
5.27. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>media-baja</i> ” del curso académico 1516.	44
5.28. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>media-baja</i> ” del curso académico 1718.	45
5.29. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>media-baja</i> ” del curso académico 1920.	46
5.30. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>media-alta</i> ” del curso académico 1516.	47
5.31. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>media-alta</i> ” del curso académico 1617.	48
5.32. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>media-alta</i> ” del curso académico 1718.	49
5.33. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>media-alta</i> ” del curso académico 1920.	50

5.34. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>alta</i> ” del curso académico 1516.	51
5.35. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>alta</i> ” del curso académico 1617.	52
5.36. Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “ <i>alta</i> ” del curso académico 1920.	53

Índice de cuadros

5.1. Muestra del primer dataset	15
5.2. Resumen del primer dataset que tenemos	15
5.3. Muestra del segundo dataset	16
5.4. Resumen del segundo dataset que tenemos	16
5.5. Resultados del test two way ANOVA	17
5.6. Coeficientes del modelo	17
5.7. Dataset 2 tras ser preparado para hacer join	19
5.8. Dataset 2 tras agrupar por grupo y año	19
5.9. Unión de ambos datasets	20
5.10. Muestra del dataset de los grupos suspensos . Consta de 2397 registros con 7 campos cada uno.	20
5.11. Muestra del dataset de los grupos aprobados . Consta de 7053 registros con 7 campos cada uno.	20
5.12. Muestra del dataset de los grupos con notable . Consta de 18840 registros con 7 campos cada uno.	21
5.13. Muestra del dataset de los grupos con sobresaliente . Consta de 17891 registros con 7 campos cada uno.	21
5.14. Muestra del dataset de los grupos con matrícula de honor . Consta de 1646 registros con 7 campos cada uno.	22

Capítulo 1

Introducción, motivaciones, objetivos y estructura

1.1. Introducción

La necesidad de comprender el proceso de aprendizaje y de personalizar la enseñanza para realizar una mejor adaptación a las necesidades del individuo ha motivado la *Analítica de Aprendizaje* o *Learning Analytics*, disciplina que consiste en la recogida de datos de un entorno de aprendizaje y el análisis de los mismos cuyo objetivo es asistir en el proceso de aprendizaje del alumnado.

Además, el uso de laboratorios virtuales y remotos en la enseñanza está en auge. Entre muchas de sus ventajas tenemos una mayor privacidad para el alumnado, accesos planificados a los mismos o soporte para reportar la actividad de los alumnos y la calificación de los mismos.

En este trabajo fin de grado se usarán datos de cinco cursos académicos obtenidos en el laboratorio virtual para sistemas multiagente de la asignatura del cuarto curso académico Desarrollo Basado en Agentes del grado de Ingeniería Informática de la Universidad de Granada (España).

El laboratorio virtual diseñado para la asignatura recoge el trabajo diario de los alumnos almacenando las interacción entre los diferentes agentes y obteniendo así un extenso dataset que nos proporciona una base sólida para el uso de diversas analíticas de aprendizaje.

Así pues, se empleará un enfoque “*data-driven*” o *impulsado por datos*, tomando decisiones estratégicas basándose en el análisis de los datos y en la interpretación de los mismos.

1.2. Motivaciones

El buen maestro hace que el mal estudiante se convierta en bueno y el buen estudiante en superior.

Maruja Torres

Como decía Maruja Torres, un buen profesor puede influenciar positivamente en el desarrollo académico de sus alumnos. Así pues, uno de las principales motivaciones de este trabajo es, precisamente, ayudar a identificar buenas y malas prácticas de los alumnos precozmente para que el profesorado pueda asistirlos mejor durante su proceso de aprendizaje y mejorar su rendimiento académico.

Así pues, la motivación última de los estudios expuesto en este documento es ayudar a aprender al alumnado.

Asimismo, creo que es de gran interés práctico conocer cuáles son las mejores estrategias a la hora de programar un sistema multiagente que desconoce el entorno que le rodea y que intenta conseguir un objetivo preestablecido.

1.3. Objetivos

Tras un primer análisis de los datos, se pretende tratar de relevar posibles estrategias escondidas en los mismos utilizando técnicas de minería de procesos, donde se considerará que una estrategia es el proceso seguido por los alumnos hasta que llegan a su objetivo.

Es decir, se pretende identificar patrones de comportamiento de los drones en los diferentes mundos virtuales con la finalidad de ayudar al profesorado de la asignatura a guiar al alumnado.

1.4. Organización de este Trabajo Fin de Grado

Este trabajo fin de grado consta de cuatro partes, siete capítulos y otros elementos como la portada, la autorización para su ubicación en la biblioteca de la escuela, sendos resúmenes tanto en español como en inglés (con sus respectivas palabras clave), la sección de agradecimientos, los índices general, de figuras y de cuadros así como una bibliografía, un glosario de términos y un glosario de acrónimos.

A continuación se expone un breve esquema general del contenido de las partes y capítulos de este trabajo fin de grado:

- Parte I:

- Parte II:
- Parte III:
- Parte IV:

Capítulo 2

Planificación

2.1. Herramientas

En la realización de un proyecto investigación, la elección de las herramientas para su desarrollo es clave. A continuación se expone una lista con las que se han utilizado este trabajo fin de grado:

- En el desarrollo del software se han empleado los lenguajes de programación `R` y `C++`.
- Se ha empleado la herramienta de minería de procesos `DISCO`.
- Se ha utilizado `GitHub` para alojar todo el contenido del proyecto y gestionar las distintas versiones del mismo.
- Para la redacción de la memoria se ha empleado el editor gratuito `Texmaker` utilizando el sistema de composición de textos `LATEX`.
- Para el seguimiento del proyecto se han utilizado las `Hojas de cálculo de Google`.

Capítulo 3

Análisis

3.1. Algunos conceptos relevantes

En el ámbito de las matemáticas y las ciencias de la computación, se emplea el término “*grafo*” (del griego “*grafos*” que significa “*dibujo*” o “*imagen*”) para referirse a un conjunto de objetos llamados “*vértices*” o “*nodos*”, los cuales están unidos por enlaces conocidos como “*aristas*” o “*arcos*”. Estas conexiones representan las relaciones binarias que existen entre los elementos de un conjunto, y son objeto de estudio de la teoría de grafos.

En la Figura 3.1 se muestra un ejemplo de grafo.

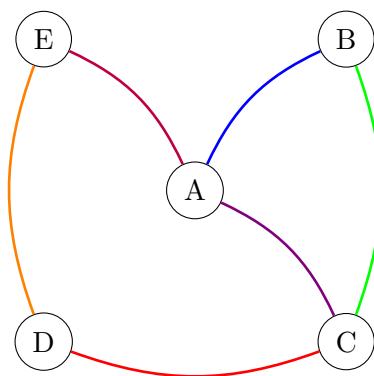


Figura 3.1: Ejemplo de grafo con 5 nodos.

Matemáticamente, un grafo $G = (V, E)$ es una tupla de vértices V y aristas E que relacionan dichos vértices. Denominaremos *orden* del grafo al número de vértices del mismo ($|V|$). Por supuesto, siempre tendremos que $V \neq \emptyset$.

Asimismo, se denomina *grafo dirigido* o *digrafo* a aquellos grafos cuyas

aristas tengan un sentido definido. Por el contrario, las aristas de los grafos no dirigidos representan relaciones simétricas.

En la Figura 3.2 se muestra un ejemplo de grafo dirigido.

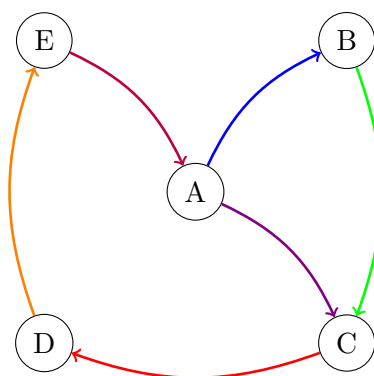


Figura 3.2: Ejemplo de grafo dirigido con 5 nodos.

Mientras que en un grafo no dirigido se tiene que $E \subseteq \{x \in \mathcal{P}(V) : |x| = 2\}$ (es decir, E es un conjunto de pares no ordenados de elementos de V), cuando el grafo es dirigido se tiene que E es un conjunto de pares ordenados $(i, j) \in V \times V$.

No obstante, en este trabajo fin de grado nos serán de gran utilidad los llamados *grafos dirigidos acíclicos* o *DAG* (*Directed Acyclic Graphs*, en inglés) que no son más que grafos dirigidos desprovistos de ciclos.

En la Figura 3.3 se muestra un ejemplo de grafo acíclico dirigido. Por el contrario, el grafo que se muestra en la figura 3.2 no es acíclico (tiene ciclos, por ejemplo, el $A-C-D-E$).

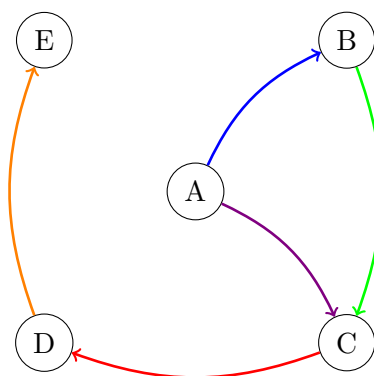


Figura 3.3: Ejemplo de grafo acíclico dirigido con 5 nodos.

3.2. Planteamiento del problema

El servidor contiene varios mundos virtuales. Cada mundo virtual es una matriz cuadrada que representa espacios abiertos (en color blanco), obstáculos (en negro) y objetivos (en rojo) tal y como se muestra en la Figura ???. Los agentes de los alumnos deben entrar en uno de esos mundos virtuales, percibir su vecindario, navegar a través de los espacios abiertos (empleando alguna clase de heurística exploratoria), evitar obstáculos y tratar de llegar al objetivo.

La percepción del agente de su entorno es crítica para resolver estos mundos. En este laboratorio virtual los alumnos pueden configurar cuál de los siguientes sensores estarán enchufados en sus agentes (cualquier combinación de ellos):

- Un **GPS** que indica al agente sus coordenadas (x, y) en el mundo virtual.
- Un **sensor de batería**. Cada agente está alimentado con una batería cuya capacidad es limitada y cuya carga decrece conforme el agente realiza algún movimiento. La batería nunca debe ser vaciada por completo.
- Un **sensor radar** que informa al agente acerca de los tipos de celdas que lo rodean con una percepción local de 5×5 (observar Figura ??).
- Un **sensor escáner** que actúa como *detector del objetivo* e indica al agente la distancia al objetivo medida desde cada una de las celdas de su entorno 5×5 (observar Figura ??).

Basados en su percepción del mundo virtual, cada agente decidirá ejecutar alguna de las siguientes acciones en su entorno implementando cualquier heurística o proceso de búsqueda.

- **LOGIN**. Entrar en cualquiera de los mundos virtuales.
- **MOVE**. Mover al agente a una de las 8 celdas adyacentes y gastar una cierta cantidad de batería. Si la celda destino es un obstáculo o el agente se queda sin batería, el agente se rompe y se sale del mundo virtual.
- **REFUEL**. El agente recarga completamente su batería. A los agentes se les permite recargar su batería tantas veces como deseen.

Capítulo 4

Diseño

Capítulo 5

Implementación

5.1. Análisis ANOVA y primeras conclusiones

5.1.1. Descripción del dataset

Los datasets que se van a usar han sido recopilados por Luis Castillo Vidal y corresponden a la actividad de sus alumnos en la asignatura [Desarrollo Basado en Agentes](#).

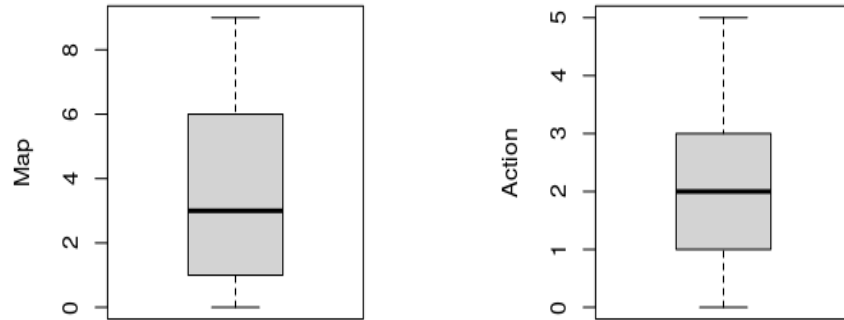
El primer dataset, tras haber sido filtrados los registros erróneos, consta de 47828 filas correspondientes a los diferentes acciones de unos drones en una serie de mundos virtuales. En cada registro se detallan los siguientes atributos:

- *year*: identifica el curso académico en el que se realizó dicha acción.
- *group*: grupo de prácticas que ha progradado al dron que acomete la acción.
- *date*: fecha en la que se lleva a cabo la acción.
- *map*: mundo virtual en el que se ha realizado la acción.
- *action*: indica el tipo de acción realizada.

En la Tabla 5.1 se presentan los primeros seis registros del dataset. Además, en la Tabla 5.2 puede apreciarse un resumen de los datos que tenemos.

La Figuras 5.1a y 5.1b muestran, respectivamente, los gráficos de caja y bigotes de las variables *map* y *action*.

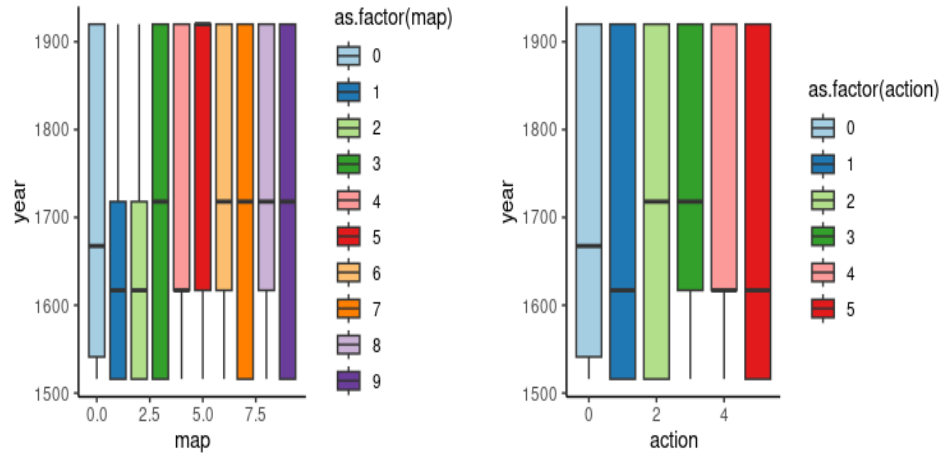
Además, podemos ver la distribución de los diferentes niveles del factor *map* y de los diferentes niveles del factor *action* en las Figuras 5.2a y 5.2b.



(a) Gráfico de caja y bigotes de la variable *map*

(b) Gráfico de caja y bigotes de la variable *action*

Figura 5.1: Diagramas de caja y bigotes de las variables *map* y *action*



(a) Distribución de los diferentes niveles del factor *map*

(b) Distribución de los diferentes niveles del factor *action*

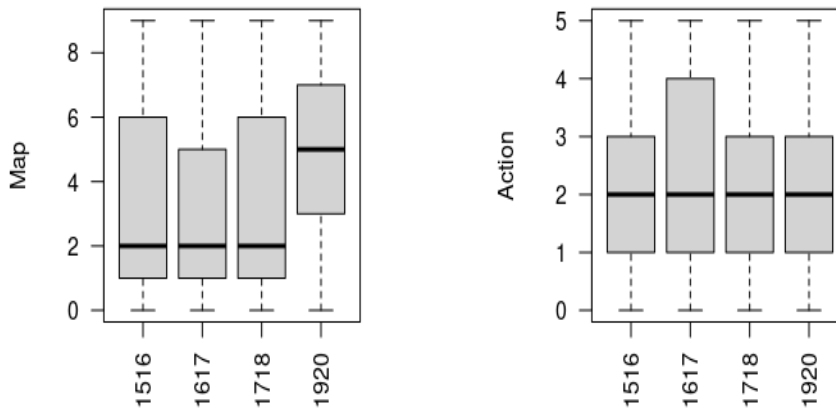
Figura 5.2: Distribuciones para diferentes niveles de los factores

	year	group	date	map	action
1	1516	Achernar	17/10/2015 19:41:45	0	0
2	1516	Bellatrix	17/10/2015 19:41:45	0	0
3	1516	Cerastes	17/10/2015 19:41:45	0	0
4	1516	Denebola	17/10/2015 19:41:45	0	0
5	1516	Elnath	17/10/2015 19:41:45	0	0
6	1516	Furud	17/10/2015 19:41:45	0	0

Cuadro 5.1: Muestra del primer dataset

year	group	date	map	action
Min. :1516	Length:47828	Length:47828	Min. :0.000	Min. :0.000
1st Qu.:1516	Class :character	Class :character	1st Qu.:1.000	1st Qu.:1.000
Median :1617	Mode :character	Mode :character	Median :3.000	Median :2.000
Mean :1700			Mean :3.834	Mean :2.325
3rd Qu.:1920			3rd Qu.:6.000	3rd Qu.:3.000
Max. :1920			Max. :9.000	Max. :5.000

Cuadro 5.2: Resumen del primer dataset que tenemos



(a) Distribución de la variable *map* en función de *year* (b) Distribución de la variable *action* en función de *year*

Figura 5.3: Distribuciones de las variables *map* y *action* dependiendo del valor de *year*

Por último, también podemos ver la distribución de las variables *map* y *action* en función del año (variable *year*) en las Figuras 5.3a y 5.3b.

El segundo dataset, tras haber sido eliminados algunas columnas que no eran interesantes para nuestro estudio, consta de 118 filas correspondientes

a las diferentes calificaciones de los equipos en las dos prácticas realizadas en la asignatura [Desarrollo Basado en Agentes](#). En cada registro se detallan los siguientes atributos:

- *Group*: grupo de prácticas que ha progradado al dron.
- *Team*: cadena de texto que identifica el curso académico en el que se realizó dicha acción, la práctica realizada y el grupo de prácticas conjuntamente.
- *Size*: tamaño del grupo de prácticas.
- *Year*: identifica el curso académico en el que se realizó dicha acción.
- *Grade*: calificación obtenida por el grupo de prácticas.

En la Tabla 5.3 se presentan los primeros seis registros del dataset. Además, en la Tabla 5.4 puede apreciarse un resumen de los datos que tenemos.

	Group	Team	Size	Year	Grade
1	G1	DBA 1819 P3 GL	4	1819	10,00
2	G2	DBA 1920 P3 GJ	4	1920	4,01
3	G3	DBA 1819 P2 GH	4	1819	7,96
4	G4	DBA 1920 P2 GE	4	1920	8,95
5	G5	DBA 1920 P3 GK	4	1920	4,51
6	G6	DBA 1415 P3 G6	6	1415	7,20

Cuadro 5.3: Muestra del segundo dataset

Group	Team	Size	Year	Grade
Length:118	Length:118	Min. :3.000	Min. :1314	Length:118
Class :character	Class :character	1st Qu.:4.000	1st Qu.:1516	Class :character
Mode :character	Mode :character	Median :5.000	Median :1718	Mode :character
		Mean :4.831	Mean :1662	
		3rd Qu.:6.000	3rd Qu.:1819	
		Max. :6.000	Max. :1920	

Cuadro 5.4: Resumen del segundo dataset que tenemos

5.1.2. Introducción

En este estudio inicial se desarrollará un modelo estadístico para determinar el efecto de los parámetros *map* y *action* (dos variables explicativas) en la variable respuesta *year*.

La relevancia de cada una de las variables en el modelo se determinará por el test *two way ANOVA* con un 5 % de nivel de significancia y se empleará la técnica de los *mínimos cuadrados* para estimar los coeficientes del modelo considerado.

5.1.3. Two way ANOVA

Un resumen de los resultados obtenidos al realizar el test two way ANOVA se muestra en la Tabla 5.5. Puede observarse que la variable *map* es significativa al nivel 0, que la variable *action* es significativa al nivel 0.01 y que la variable *map:action* (el término de interacción) no es significativa. Así pues, puede concluirse que el dataset es homogéneo, es decir, las combinaciones *map:action* son estadísticamente iguales en todos los años considerados.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
map	1	62918413.58	62918413.58	2689.15	0.0000
action	1	101244.69	101244.69	4.33	0.0375
map:action	1	8797.03	8797.03	0.38	0.5398
Residuals	47824	1118946173.13	23397.17		

Cuadro 5.5: Resultados del test two way ANOVA

La notación escalar del modelo ajustado al aplicar el test tiene la siguiente estructura:

$$y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_1 \cdot x_2 + \epsilon \quad (5.1)$$

donde β_0 es el intercepto, β_1 y β_2 son los coeficientes de los efectos principales, β_3 es el coeficiente del término de interacción, x_1 y x_2 son los parámetros sometidos a investigación (en este caso, x_1 representa el parámetro mapa y x_2 representa la acción), y representa el año y ϵ es el *término error*.

La Tabla 5.6 muestra los valores de los coeficientes de la fórmula que se han obtenido tras ajustar el modelo de regresión a los datos.

	x
(Intercept)	1655.03
map	12.31
action	-1.51
map:action	0.12

Cuadro 5.6: Coeficientes del modelo

La Figura 5.4 muestra la interacción entre los parámetros mapa y acción.

Así pues, puede observarse que todas las líneas de la gráfica siguen más o menos el mismo patrón, lo que evidencia que no hay una gran interacción entre ambos.

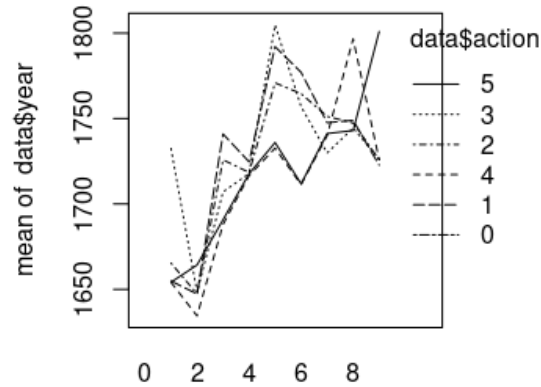


Figura 5.4: Interacción entre las variables *map* y *action*

La Figura 5.5 muestra que no se violan las suposiciones que hemos realizado sobre el modelo. La media y la varianza de los residuos no parece que varíe respecto de los valores ajustados. Como consecuencia, concluiré que podemos suponer la homocedasticidad. Además, si nos fijamos en el *Normal Q-Q plot*, puede observarse que los residuos son gaussianos.

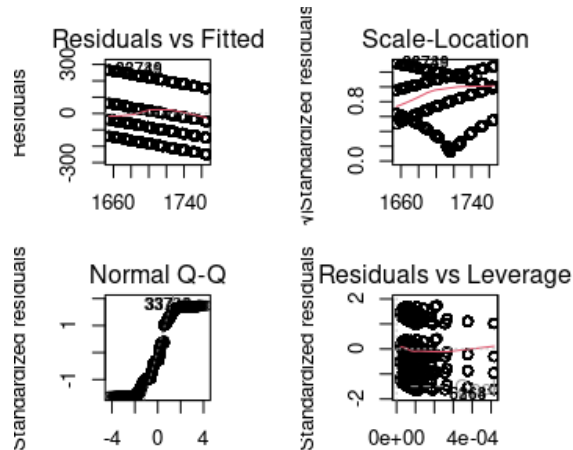


Figura 5.5: Gráficas diagnósticas del modelo ANOVA

5.1.4. Segmentación de los datos

Se ha realizado una segmentación de los registros del primer dataset en función de la calificación obtenida por los grupos que han realizado las acciones en los mapas.

Para ello, se ha sustituido la columna *Team* del segundo dataset por el nombre del grupo en cada caso y se ha eliminado la columna *Group*. Después de realizar este proceso el dataset tiene la forma que se muestra en la Tabla 5.7

	group	size	year	grade
1	Lesath	4	1819	10,00
2	Jabbah	4	1920	4,01
3	Haldus	4	1819	7,96
4	Elnath	4	1920	8,95
5	Keid	4	1920	4,51
6	Furud	6	1415	7,20

Cuadro 5.7: Dataset 2 tras ser preparado para hacer join

A continuación, se transformarán los valores de la columna *grade* a tipo `double` y se agruparán los datos de este dataset por grupo y año, calculando la calificación media de las prácticas para cada par grupo-año. El resultado de realizar esta operación puede apreciarse en la Tabla 5.8.

	group	year	size	mean_grade
1	Achernar	1314	6	9.40
2	Achernar	1415	5	9.00
3	Achernar	1516	5	8.20
4	Achernar	1617	5	10.00
5	Achernar	1718	5	8.33
6	Bellatrix	1516	5	8.20

Cuadro 5.8: Dataset 2 tras agrupar por grupo y año

Ahora, se juntan ambos datasets usando la función `inner_join` por grupo y año. El resultado de esta operación puede observarse en la Tabla 5.9.

Por último, se separarán los datos de esta última tabla en función de las calificaciones obtenidas por los diferentes grupos:

- **Suspense:** calificación mayor o igual que 0 y menor que 5.
- **Aprobado:** calificación mayor o igual que 5 y menor que 7.
- **Notable:** calificación mayor o igual que 7 y menor que 9.

	group	year	size	mean_grade	date	map	action
1	Achernar	1516	5	8.2	17/10/2015 19:41:45	0	0
2	Achernar	1516	5	8.2	22/10/2015 17:29:21	1	1
3	Achernar	1516	5	8.2	22/10/2015 17:29:22	1	2
4	Achernar	1516	5	8.2	22/10/2015 17:29:39	1	3
5	Achernar	1516	5	8.2	22/10/2015 17:34:09	1	1
6	Achernar	1516	5	8.2	22/10/2015 17:34:10	1	2

Cuadro 5.9: Unión de ambos datasets

- **Sobresaliente:** calificación mayor o igual que 9 y menor que 10.
- **Matrícula de Honor:** calificación igual a 10.

Una muestra de los datasets generados tras la segmentación puede apreciarse en las tablas 5.10, 5.11, 5.12, 5.13 y 5.14.

	group	year	size	mean_grade	date	map	action
1	Girtab	1617	6	4.67	15/11/2016 7:31:31	1	1
2	Girtab	1617	6	4.67	15/11/2016 7:31:53	1	1
3	Girtab	1617	6	4.67	15/11/2016 12:50:03	1	1
4	Girtab	1617	6	4.67	15/11/2016 12:50:28	1	1
5	Girtab	1617	6	4.67	15/11/2016 13:27:27	1	1
6	Girtab	1617	6	4.67	15/11/2016 13:27:58	1	1

Cuadro 5.10: Muestra del dataset de los grupos **suspensos**. Consta de 2397 registros con 7 campos cada uno.

	group	year	size	mean_grade	date	map	action
1	Bellatrix	1920	4	6.835	05/11/2019 10:24:51	1	1
2	Bellatrix	1920	4	6.835	05/11/2019 10:24:51	1	2
3	Bellatrix	1920	4	6.835	05/11/2019 10:25:04	1	1
4	Bellatrix	1920	4	6.835	05/11/2019 10:25:26	1	2
5	Bellatrix	1920	4	6.835	05/11/2019 10:25:39	1	1
6	Bellatrix	1920	4	6.835	05/11/2019 10:25:45	1	2

Cuadro 5.11: Muestra del dataset de los grupos **aprobados**. Consta de 7053 registros con 7 campos cada uno.

	group	year	size	mean_grade	date	map	action
1	Achernar	1516	5	8.2	17/10/2015 19:41:45	0	0
2	Achernar	1516	5	8.2	22/10/2015 17:29:21	1	1
3	Achernar	1516	5	8.2	22/10/2015 17:29:22	1	2
4	Achernar	1516	5	8.2	22/10/2015 17:29:39	1	3
5	Achernar	1516	5	8.2	22/10/2015 17:34:09	1	1
6	Achernar	1516	5	8.2	22/10/2015 17:34:10	1	2

Cuadro 5.12: Muestra del dataset de los grupos con **notable**. Consta de 18840 registros con 7 campos cada uno.

	group	year	size	mean_grade	date	map	action
1	Bellatrix	1617	5	9.72	07/11/2016 11:13:31	1	1
2	Bellatrix	1617	5	9.72	07/11/2016 11:36:59	1	1
3	Bellatrix	1617	5	9.72	07/11/2016 11:41:07	1	1
4	Bellatrix	1617	5	9.72	07/11/2016 11:42:22	1	1
5	Bellatrix	1617	5	9.72	07/11/2016 11:45:37	1	1
6	Bellatrix	1617	5	9.72	07/11/2016 11:49:23	1	1

Cuadro 5.13: Muestra del dataset de los grupos con **sobresaliente**. Consta de 17891 registros con 7 campos cada uno.

5.2. Extracción de procesos ocultos en los datasets

5.2.1. Extracción de los procesos con DISCO

Del dataset proporcionado se extraerán los campos de información más importantes:

1. El identificador del caso, extraído de una clave aleatoria generada al principio de cada operación **LOGIN** y que distingue de manera unívoca cada sesión de trabajo de los estudiantes.
2. El agente, que se refiere al nombre del grupo de estudiantes.
3. La fecha y hora a la que se registró la transacción.
4. El campo actividad (**Activity**), que refleja la acción de los alumnos en el mundo virtual.
5. Varios campos de tipo recurso (**Resource**) que proporcionan información adicional que puede sernos de utilidad a la hora de filtrar los registros.

Se importarán el dataset de dos maneras diferentes, con el objetivo de estudiar tanto la frecuencia con que cada problema o mapa ha sido visitado

	group	year	size	mean_grade	date	map	action
1	Achernar	1617	5	10	09/11/2016 21:23:22	1	1
2	Achernar	1617	5	10	09/11/2016 21:25:11	1	1
3	Achernar	1617	5	10	10/11/2016 0:21:37	1	1
4	Achernar	1617	5	10	10/11/2016 0:22:50	1	1
5	Achernar	1617	5	10	10/11/2016 0:23:12	1	1
6	Achernar	1617	5	10	10/11/2016 0:25:53	1	1

Cuadro 5.14: Muestra del dataset de los grupos con **matrícula de honor**. Consta de 1646 registros con 7 campos cada uno.

como las acciones compuestas mapa-porcentaje superado. En la primera importación la **Activity** es la columna *Objeto* mientras que en la segunda es la columna *Compuesto* tal y como puede verse en las Figuras 5.6 y 5.8.

	Grupo	Año	Tiempo	Session	Objeto	Accion	Nota	Compuesto	Column 9	Column 10
1	DBA 1516 P2 GA	1516	22/10/2015 17:29:21	84557317827	PROBLEM1	20	7.8	PROBLEM1-20		
2	DBA 1516 P2 GA	1516	22/10/2015 17:29:22	84557317827	PROBLEM1	40	7.8	PROBLEM1-40		
3	DBA 1516 P2 GA	1516	22/10/2015 17:29:39	84557317827	PROBLEM1	60	7.8	PROBLEM1-60		
4	DBA 1516 P2 GA	1516	22/10/2015 17:34:09	13090455475	PROBLEM1	20	7.8	PROBLEM1-20		
5	DBA 1516 P2 GA	1516	22/10/2015 17:34:10	13090455475	PROBLEM1	40	7.8	PROBLEM1-40		
6	DBA 1516 P2 GA	1516	22/10/2015 17:34:40	13090455475	PROBLEM1	60	7.8	PROBLEM1-60		
7	DBA 1516 P2 GA	1516	28/10/2015 14:08:56	103879609524	PROBLEM1	20	7.8	PROBLEM1-20		
8	DBA 1516 P2 GA	1516	02/11/2015 16:13:41	93419335161	PROBLEM1	20	7.8	PROBLEM1-20		
9	DBA 1516 P2 GA	1516	02/11/2015 16:17:25	98069190243	PROBLEM1	20	7.8	PROBLEM1-20		
10	DBA 1516 P2 GA	1516	02/11/2015 16:24:55	82317887407	PROBLEM1	20	7.8	PROBLEM1-20		
11	DBA 1516 P2 GA	1516	02/11/2015 16:25:22	50615350630	PROBLEM1	20	7.8	PROBLEM1-20		
12	DBA 1516 P2 GA	1516	02/11/2015 16:26:28	23299516035	PROBLEM1	20	7.8	PROBLEM1-20		
13	DBA 1516 P2 GA	1516	02/11/2015 16:39:44	72144600342	PROBLEM1	20	7.8	PROBLEM1-20		
14	DBA 1516 P2 GA	1516	02/11/2015 16:40:26	85388844511	PROBLEM1	20	7.8	PROBLEM1-20		
15	DBA 1516 P2 GA	1516	02/11/2015 16:43:27	16755827257	PROBLEM1	20	7.8	PROBLEM1-20		
16	DBA 1516 P2 GA	1516	02/11/2015 16:45:06	89590155452	PROBLEM1	20	7.8	PROBLEM1-20		
17	DBA 1516 P2 GA	1516	02/11/2015 16:48:28	60330564316	PROBLEM1	20	7.8	PROBLEM1-20		
18	DBA 1516 P2 GA	1516	02/11/2015 16:48:36	6008106071	PROBLEM1	20	7.8	PROBLEM1-20		
19	DBA 1516 P2 GA	1516	02/11/2015 16:50:53	109827932573	PROBLEM1	20	7.8	PROBLEM1-20		
20	DBA 1516 P2 GA	1516	02/11/2015 16:51:23	38856545905	PROBLEM1	20	7.8	PROBLEM1-20		
21	DBA 1516 P2 GA	1516	02/11/2015 16:51:45	52924525286	PROBLEM1	20	7.8	PROBLEM1-20		
22	DBA 1516 P2 GA	1516	02/11/2015 16:55:07	63999148696	PROBLEM1	20	7.8	PROBLEM1-20		
23	DBA 1516 P2 GA	1516	02/11/2015 16:56:29	78342623435	PROBLEM1	20	7.8	PROBLEM1-20		
24	DBA 1516 P2 GA	1516	02/11/2015 17:00:58	21506524143	PROBLEM1	20	7.8	PROBLEM1-20		
25	DBA 1516 P2 GA	1516	02/11/2015 17:03:00	70390370747	PROBLEM1	20	7.8	PROBLEM1-20		
26	DBA 1516 P2 GA	1516	02/11/2015 17:03:47	76509787638	PROBLEM1	20	7.8	PROBLEM1-20		
27	DBA 1516 P2 GA	1516	02/11/2015 17:04:31	73862232129	PROBLEM1	20	7.8	PROBLEM1-20		
28	DBA 1516 P2 GA	1516	02/11/2015 17:05:17	97284175030	PROBLEM1	20	7.8	PROBLEM1-20		
29	DBA 1516 P2 GA	1516	02/11/2015 17:06:45	26488905361	PROBLEM1	20	7.8	PROBLEM1-20		
30	DBA 1516 P2 GA	1516	02/11/2015 17:07:12	108197219076	PROBLEM1	20	7.8	PROBLEM1-20		
31	DBA 1516 P2 GA	1516	02/11/2015 17:08:11	28262256652	PROBLEM1	20	7.8	PROBLEM1-20		
32	DBA 1516 P2 GA	1516	02/11/2015 17:09:42	9385851428	PROBLEM1	20	7.8	PROBLEM1-20		
33	DBA 1516 P2 GA	1516	02/11/2015 17:10:13	70387634948	PROBLEM1	20	7.8	PROBLEM1-20		
34	DBA 1516 P2 GA	1516	02/11/2015 17:10:44	88589167061	PROBLEM1	20	7.8	PROBLEM1-20		
35	DBA 1516 P2 GA	1516	02/11/2015 17:44:52	99057369615	PROBLEM1	20	7.8	PROBLEM1-20		
36	DBA 1516 P2 GA	1516	02/11/2015 17:45:34	10528362411	PROBLEM1	20	7.8	PROBLEM1-20		
37	DBA 1516 P2 GA	1516	02/11/2015 17:46:37	30704231316	PROBLEM1	20	7.8	PROBLEM1-20		
38	DBA 1516 P2 GA	1516	02/11/2015 18:01:26	65467626545	PROBLEM1	20	7.8	PROBLEM1-20		
39	DBA 1516 P2 GA	1516	02/11/2015 18:02:20	64662077168	PROBLEM1	20	7.8	PROBLEM1-20		
40	DBA 1516 P2 GA	1516	02/11/2015 18:03:48	62389948304	PROBLEM1	20	7.8	PROBLEM1-20		

Figura 5.6: Análisis de procesos del dataset compuesto (acción mapa).

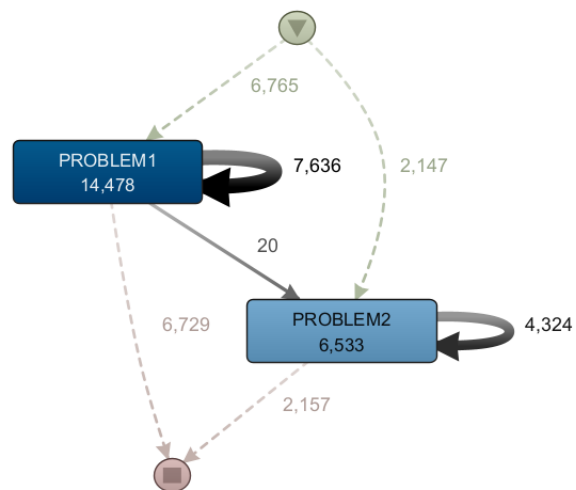


Figura 5.7: Análisis de procesos del dataset compuesto (acción mapa).

Disco - New project

Session
column is used

Case

	Grupo	Año	Tiempo	Session	Objeto	Accion	Nota	Compuesto	Column 9	Column 10
1	DBA 1516 P2 GA	1516	22/10/2015 17:29:21	84557317827	PROBLEM1	20	7.8	PROBLEM1-20		
2	DBA 1516 P2 GA	1516	22/10/2015 17:29:22	84557317827	PROBLEM1	40	7.8	PROBLEM1-40		
3	DBA 1516 P2 GA	1516	22/10/2015 17:29:39	84557317827	PROBLEM1	60	7.8	PROBLEM1-60		
4	DBA 1516 P2 GA	1516	22/10/2015 17:34:09	13090455475	PROBLEM1	20	7.8	PROBLEM1-20		
5	DBA 1516 P2 GA	1516	22/10/2015 17:34:10	13090455475	PROBLEM1	40	7.8	PROBLEM1-40		
6	DBA 1516 P2 GA	1516	22/10/2015 17:34:40	13090455475	PROBLEM1	60	7.8	PROBLEM1-60		
7	DBA 1516 P2 GA	1516	28/10/2015 14:08:56	103879609524	PROBLEM1	20	7.8	PROBLEM1-20		
8	DBA 1516 P2 GA	1516	02/11/2015 16:13:41	83419335161	PROBLEM1	20	7.8	PROBLEM1-20		
9	DBA 1516 P2 GA	1516	02/11/2015 16:17:25	88068190243	PROBLEM1	20	7.8	PROBLEM1-20		
10	DBA 1516 P2 GA	1516	02/11/2015 16:24:55	82317887407	PROBLEM1	20	7.8	PROBLEM1-20		
11	DBA 1516 P2 GA	1516	02/11/2015 16:25:22	50615350630	PROBLEM1	20	7.8	PROBLEM1-20		
12	DBA 1516 P2 GA	1516	02/11/2015 16:26:28	23299516035	PROBLEM1	20	7.8	PROBLEM1-20		
13	DBA 1516 P2 GA	1516	02/11/2015 16:39:44	72144600342	PROBLEM1	20	7.8	PROBLEM1-20		
14	DBA 1516 P2 GA	1516	02/11/2015 16:40:26	85388844511	PROBLEM1	20	7.8	PROBLEM1-20		
15	DBA 1516 P2 GA	1516	02/11/2015 16:43:27	16755827257	PROBLEM1	20	7.8	PROBLEM1-20		
16	DBA 1516 P2 GA	1516	02/11/2015 16:45:06	89890165452	PROBLEM1	20	7.8	PROBLEM1-20		
17	DBA 1516 P2 GA	1516	02/11/2015 16:48:28	60330564318	PROBLEM1	20	7.8	PROBLEM1-20		
18	DBA 1516 P2 GA	1516	02/11/2015 16:48:36	6008106071	PROBLEM1	20	7.8	PROBLEM1-20		
19	DBA 1516 P2 GA	1516	02/11/2015 16:50:53	109827932573	PROBLEM1	20	7.8	PROBLEM1-20		
20	DBA 1516 P2 GA	1516	02/11/2015 16:51:23	38856545905	PROBLEM1	20	7.8	PROBLEM1-20		
21	DBA 1516 P2 GA	1516	02/11/2015 16:51:45	52924525286	PROBLEM1	20	7.8	PROBLEM1-20		
22	DBA 1516 P2 GA	1516	02/11/2015 16:55:07	63896148696	PROBLEM1	20	7.8	PROBLEM1-20		
23	DBA 1516 P2 GA	1516	02/11/2015 16:56:29	78342623435	PROBLEM1	20	7.8	PROBLEM1-20		
24	DBA 1516 P2 GA	1516	02/11/2015 17:00:58	21506524143	PROBLEM1	20	7.8	PROBLEM1-20		
25	DBA 1516 P2 GA	1516	02/11/2015 17:03:00	70390370747	PROBLEM1	20	7.8	PROBLEM1-20		
26	DBA 1516 P2 GA	1516	02/11/2015 17:03:47	76509787638	PROBLEM1	20	7.8	PROBLEM1-20		
27	DBA 1516 P2 GA	1516	02/11/2015 17:04:31	73862232129	PROBLEM1	20	7.8	PROBLEM1-20		
28	DBA 1516 P2 GA	1516	02/11/2015 17:05:17	97284175030	PROBLEM1	20	7.8	PROBLEM1-20		
29	DBA 1516 P2 GA	1516	02/11/2015 17:06:45	26486905361	PROBLEM1	20	7.8	PROBLEM1-20		
30	DBA 1516 P2 GA	1516	02/11/2015 17:07:12	108197219076	PROBLEM1	20	7.8	PROBLEM1-20		
31	DBA 1516 P2 GA	1516	02/11/2015 17:08:11	28232250652	PROBLEM1	20	7.8	PROBLEM1-20		
32	DBA 1516 P2 GA	1516	02/11/2015 17:09:42	93865861428	PROBLEM1	20	7.8	PROBLEM1-20		
33	DBA 1516 P2 GA	1516	02/11/2015 17:10:13	70387634048	PROBLEM1	20	7.8	PROBLEM1-20		
34	DBA 1516 P2 GA	1516	02/11/2015 17:10:44	88588167061	PROBLEM1	20	7.8	PROBLEM1-20		
35	DBA 1516 P2 GA	1516	02/11/2015 17:44:52	99057369615	PROBLEM1	20	7.8	PROBLEM1-20		
36	DBA 1516 P2 GA	1516	02/11/2015 17:45:34	10528362411	PROBLEM1	20	7.8	PROBLEM1-20		
37	DBA 1516 P2 GA	1516	02/11/2015 17:46:37	30704231316	PROBLEM1	20	7.8	PROBLEM1-20		
38	DBA 1516 P2 GA	1516	02/11/2015 18:01:26	65467626545	PROBLEM1	20	7.8	PROBLEM1-20		
39	DBA 1516 P2 GA	1516	02/11/2015 18:02:20	64662077165	PROBLEM1	20	7.8	PROBLEM1-20		
40	DBA 1516 P2 GA	1516	02/11/2015 18:03:48	62389948304	PROBLEM1	20	7.8	PROBLEM1-20		

Cancel File encoding: UTF-8 Use quotes

Figura 5.8: Análisis de procesos del dataset compuesto (acción compuesta).

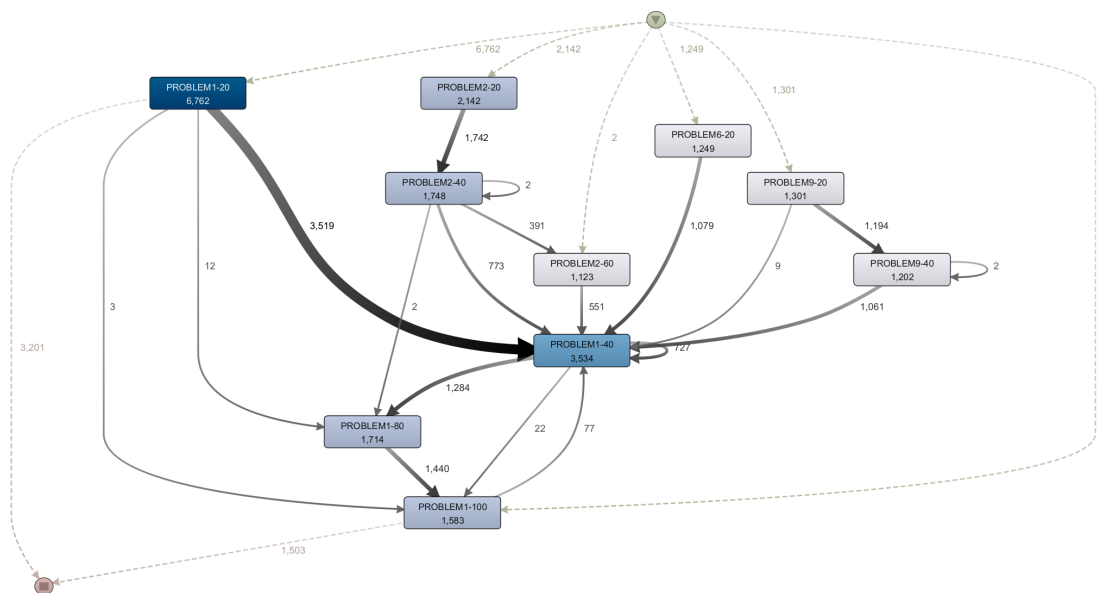
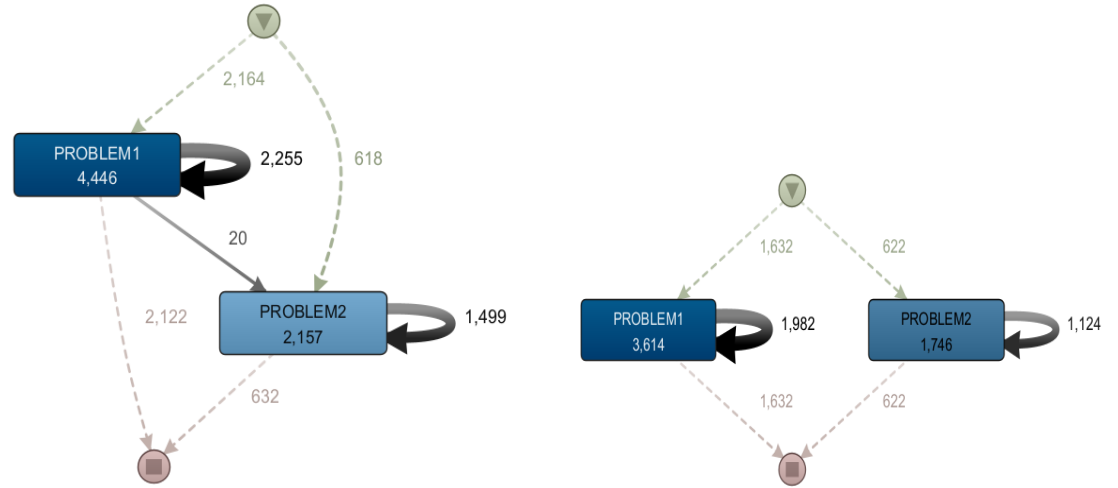
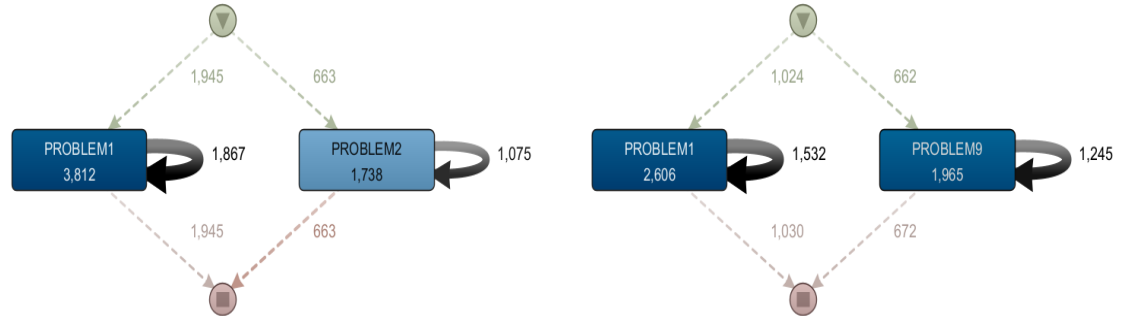


Figura 5.9: Análisis de procesos del dataset compuesto (acción compuesta).

5.2.2. Segmentación por años



(a) Extracción de procesos del curso académico 1516 (acción mapa). (b) Extracción de procesos del curso académico 1617 (acción mapa).



(c) Extracción de procesos del curso académico 1718 (acción mapa). (d) Extracción de procesos del curso académico 1920 (acción mapa).

Figura 5.10: Extracción de procesos de los diferentes cursos académicos (acción mapa/problema).

Como puede observarse, mientras que en los cursos académicos 1516, 1617 y 1718 predominan los problemas 1 y 2, en el curso académico lo hacen

los problemas 1 y 9. Esto refleja que los alumnos del curso escolar 1920 se tomaron gran interés en intentar resolver el noveno y más difícil de los problemas.

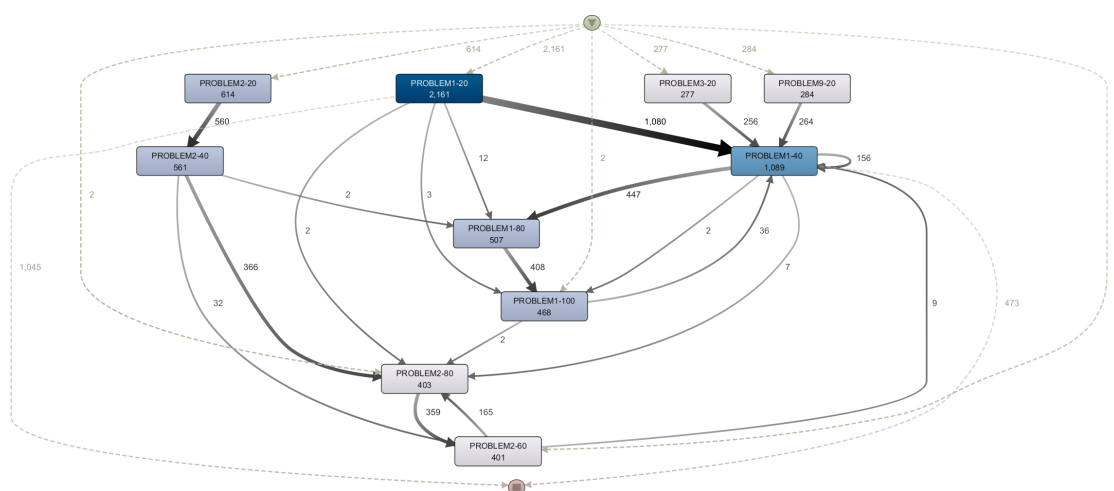


Figura 5.11: Extracción de procesos del curso académico 1516 (acción compuesta).

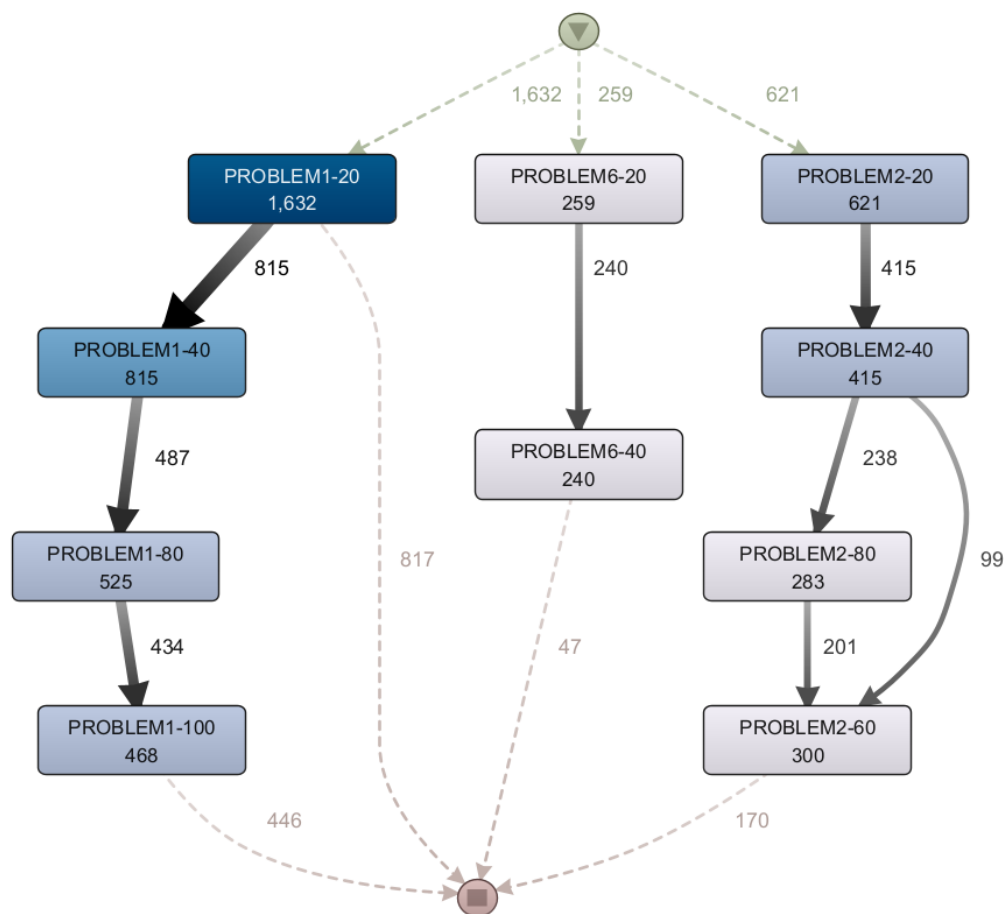


Figura 5.12: Extracción de procesos del curso académico 1617 (acción compuesta).

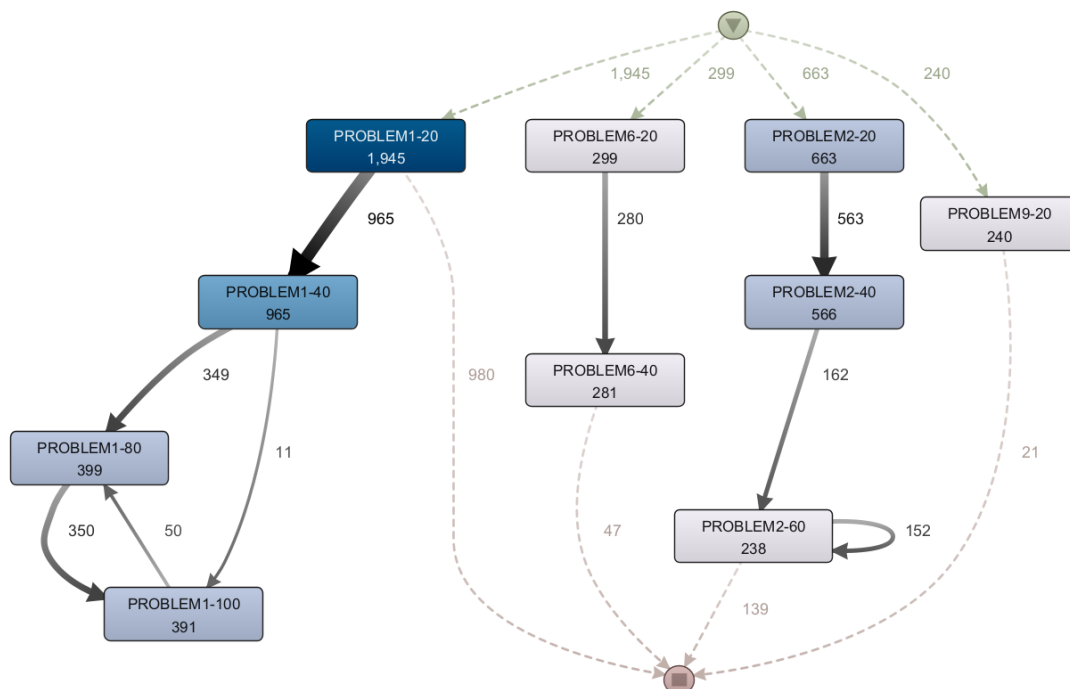


Figura 5.13: Extracción de procesos del curso académico 1718 (acción compuesta).

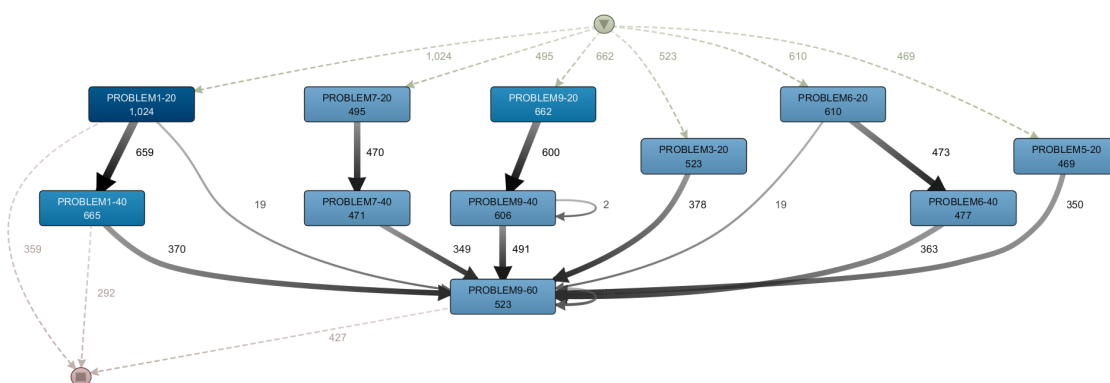
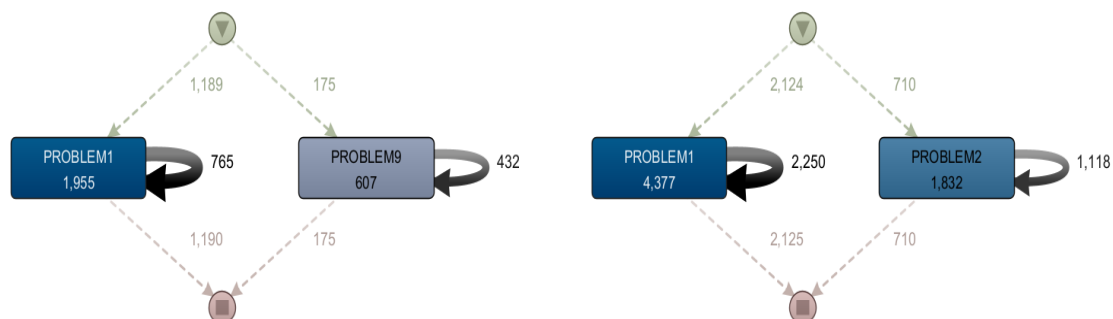


Figura 5.14: Extracción de procesos del curso académico 1920 (acción compuesta).

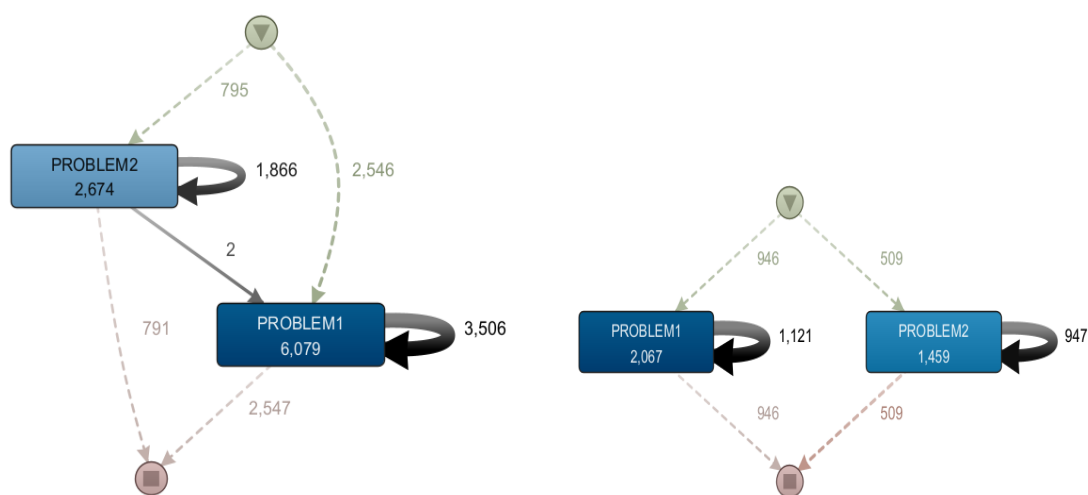
Como podemos observar en la Figura 5.14, los procesos extraídos en el curso académico 1920 resultan ser algo distintos a los del resto de cursos, pues parece que hay un mayor variedad, habiendo nodos de hasta seis problemas diferentes y estando estos más balanceados.

5.2.3. Segmentación por calificaciones

De ahora en adelante se considerará que la calificación de un grupo es “*baja*” si es inferior a 7.5, “*media-baja*” si es igual o mayor que 7.5 y menor a 8.5, “*media-alta*” si es igual a o mayor que 8.5 e inferior a 9.5 y “*alta*” si es igual o superior a 9.5.



(a) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “baja”. (b) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “media-baja”.



(c) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “media-alta”. (d) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “alta”.

Figura 5.15: Extracción de procesos del dataset integrado por las acciones mapa de los grupos según sus calificaciones.

Tras haber extraído los correspondientes grafos, notamos que los grupos con calificación “baja” han frecuentado más los problemas 1 y 9 mientras que los problemas más visitados por el resto de grupos son el 1 y el 2. Esto puede

sugerir los grupos con “peor” calificación no han frecuentado lo suficiente los problemas de menor dificultad y, por el contrario, han intentado resolver la más difícil de las pruebas.

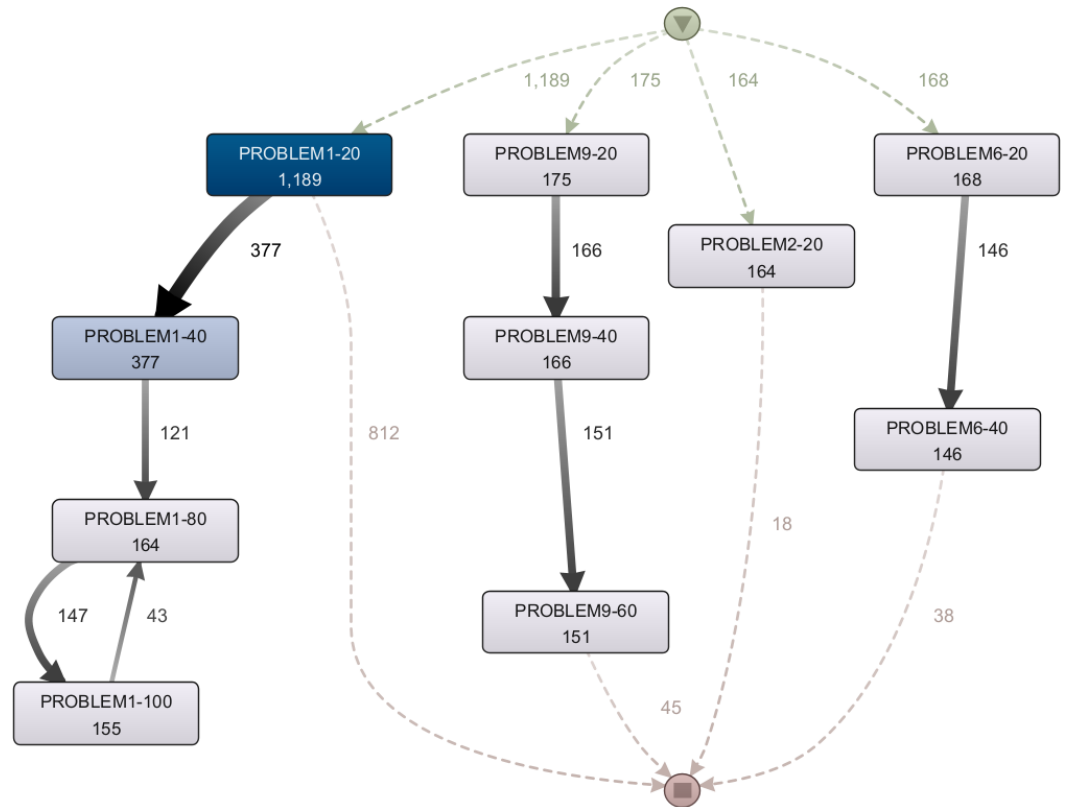


Figura 5.16: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “baja”.

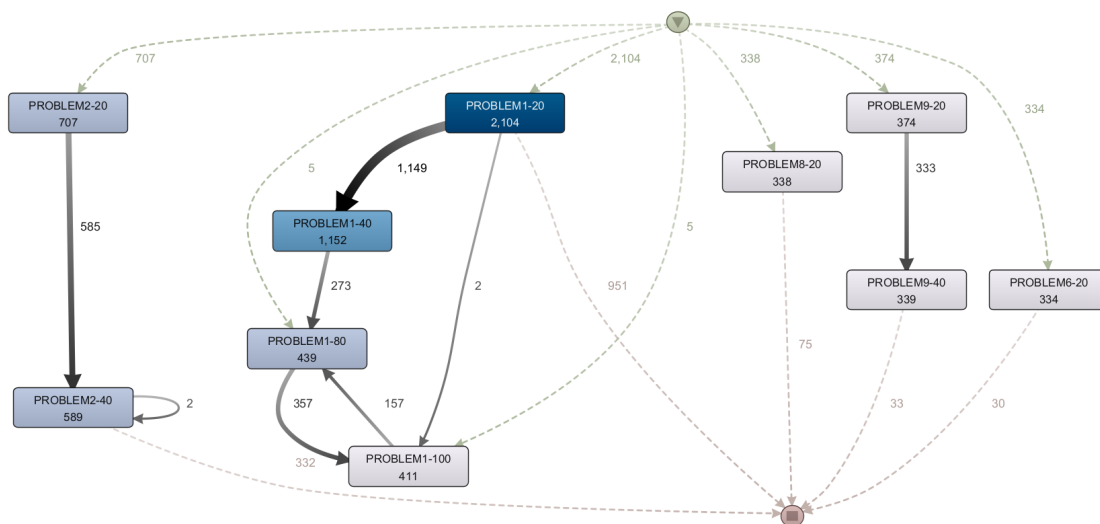


Figura 5.17: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación "media-baja".

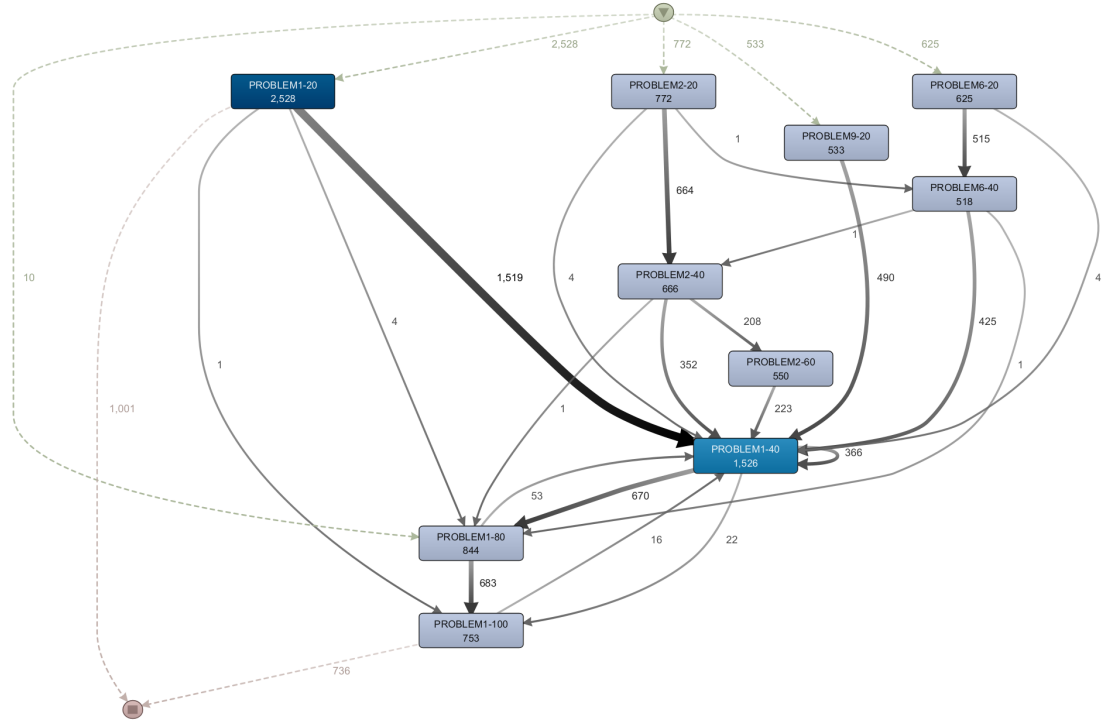


Figura 5.18: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación "media-alta".

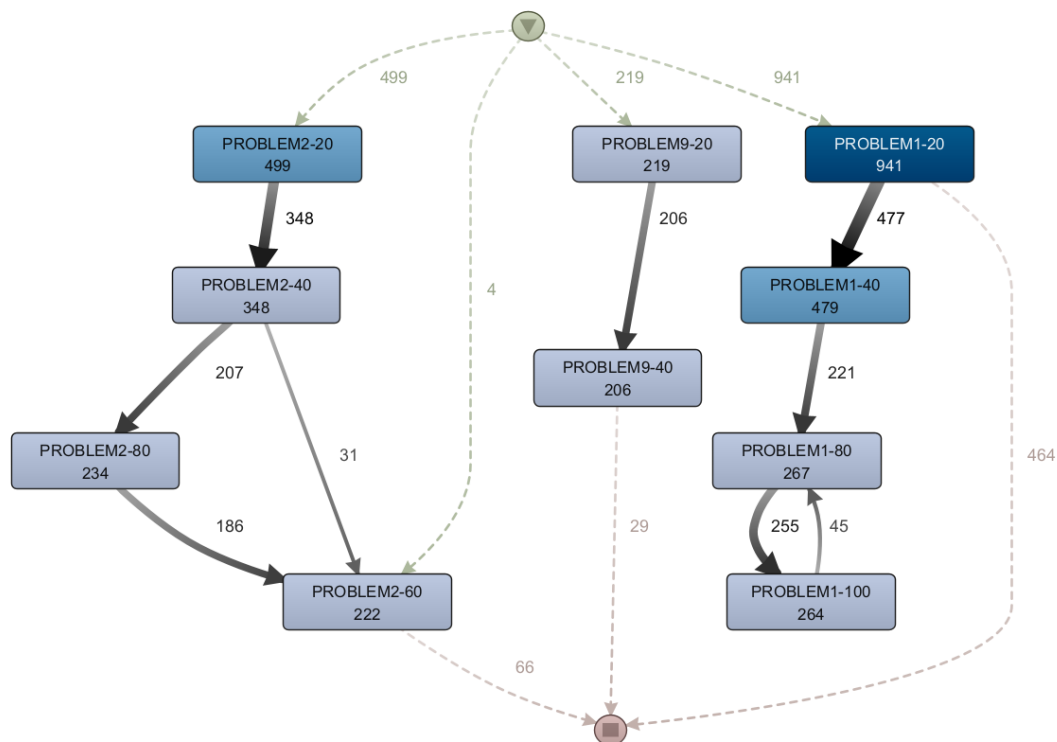
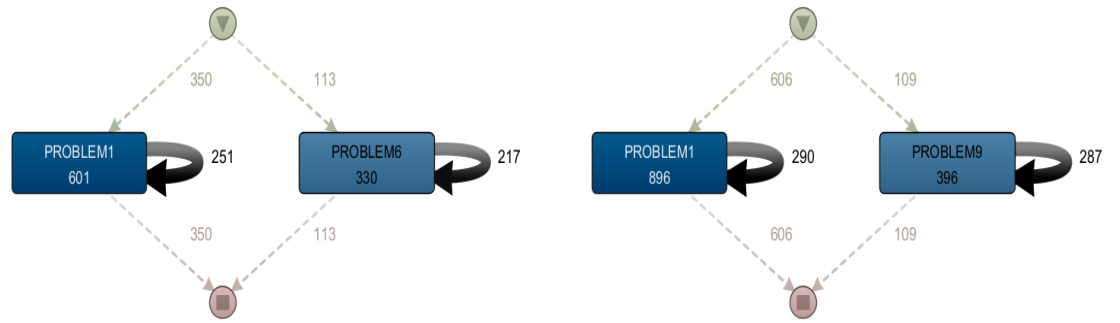


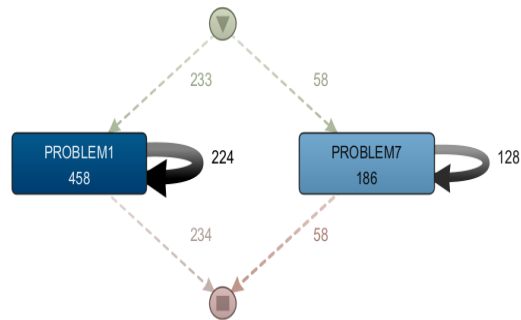
Figura 5.19: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “alta”.

Observando las Figuras 5.16, 5.17, 5.18 y 5.19, podemos ver que hay una correlación directa de la nota que sacan con la frecuencia relativa del problema con la que frecuenta el nodo PROBLEM2-20.

5.2.4. Segmentación por año y calificación



(a) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “baja” del curso académico 1617. (b) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “baja” del curso académico 1718.

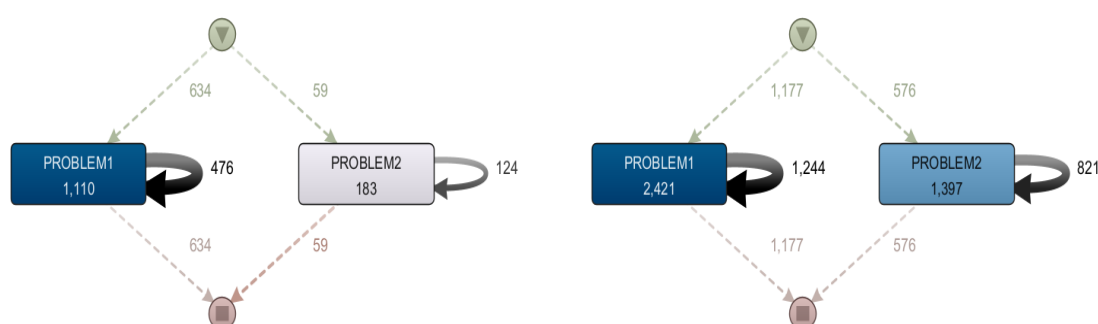


(c) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “baja” del curso académico 1920.

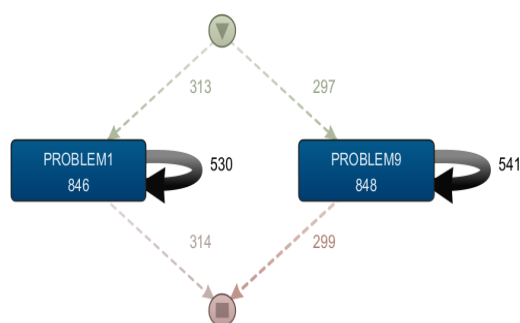
Figura 5.20: Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “baja” a lo largo de cuatro cursos académicos.

Tal y como puede observarse en la Figura que se muestra anteriormente, los grupos con las peores calificaciones a lo largo de los cuatro cursos académicos tienden a emplear mucho tiempo en el problema 1 y en otro problema con una numeración alta (en este caso, en el curso 1617 es el pro-

blema 6, en el curso 1718 es el problema 9 y en el 1920 es el problema 7). Esto podría indicar que no prestan demasiada atención a resolver problemas más fáciles, intentando resolver los más difíciles a lo mejor sin los conocimientos o la base necesaria. No obstante, también podría haber ocurrido que esos problemas en particular hubieran sido un quebradero de cabeza para esos grupos en esos cursos. Así pues, podrían haberse quedado atascados y haber invertido muchos más recursos en esos problemas que en el resto.



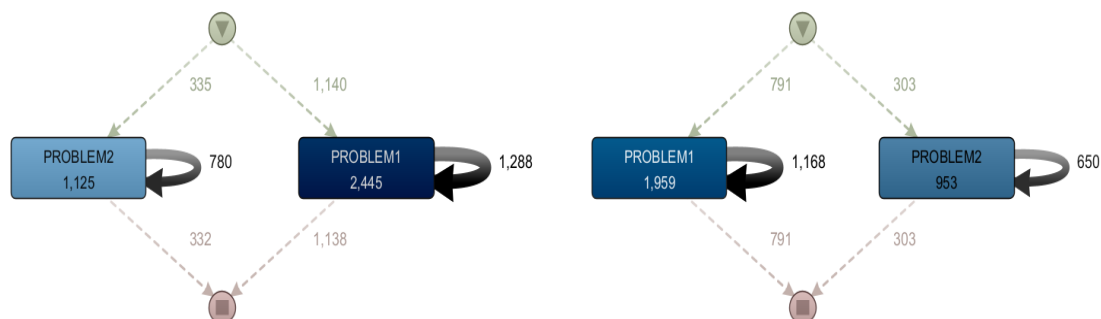
(a) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “media-baja” del curso académico 1516. (b) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “media-baja” del curso académico 1718.



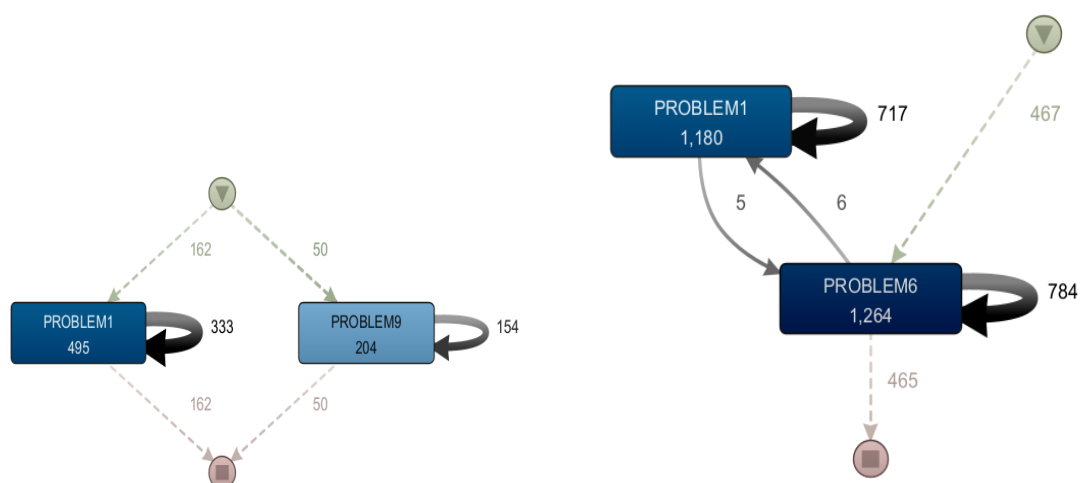
(c) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “media-baja” del curso académico 1920.

Figura 5.21: Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “media-baja” a lo largo de cuatro cursos académicos.

Podemos observar que hay claras diferencias con respecto a la Figura anterior. Por ejemplo, en los cursos académicos 1516 y 1718 los problemas de numeración más baja (los problemas 1 y 2) sobresalen con las frecuencias más altas. Esto puede indicar que a lo mejor han resuelto los diferentes de manera más progresiva, sin lanzarse a resolver los problemas más difíciles desde el principio. Sin embargo, en el curso académico 1920 todavía podemos observar aún lo mencionado en los grafos de los procesos extraídos de los grupos de alumnos con calificación *“baja”*.



(a) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “media-alta” del curso académico 1516. (b) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “media-alta” del curso académico 1617.

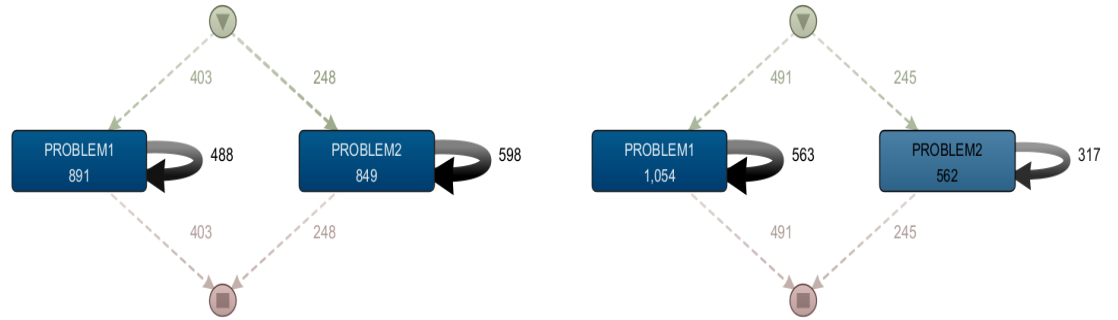


(c) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “media-alta” del curso académico 1718. (d) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “media-alta” del curso académico 1920.

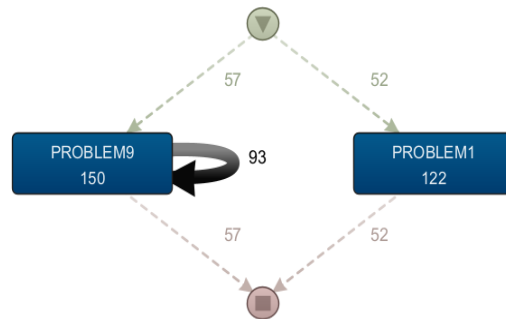
Figura 5.22: Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “media-alta” a lo largo de cuatro cursos académicos.

En los grupos con calificación “media-alta” podemos observar dos comportamientos diferentes. Por un lado, en los primeros cursos escolares (1617

y 1617) se sigue la tendencia de la Figura anterior, con la frecuencia de los problemas 1 y 2 predominando sobre las demás. Sin embargo, en el curso académico 1718, el segundo problema más frecuentado es el 9 y en el curso académico 1920 el 6 (siendo el problema 1 el más visitado en ambos casos).



(a) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “alta” del curso académico 1516. (b) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “alta” del curso académico 1617.



(c) Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “alta” del curso académico 1920.

Figura 5.23: Extracción de procesos del dataset integrado por las acciones mapa de los grupos con calificación “alta” a lo largo de cuatro cursos académicos.

El mismo patrón puede observarse en los grafos de los grupos con las calificaciones más altas en los cuatro cursos académicos, siendo los problemas

1 y 9 los más frecuentados en el curso 1920, mientras en el resto de los cursos son los problemas 1 y 2 los más frecuentados.

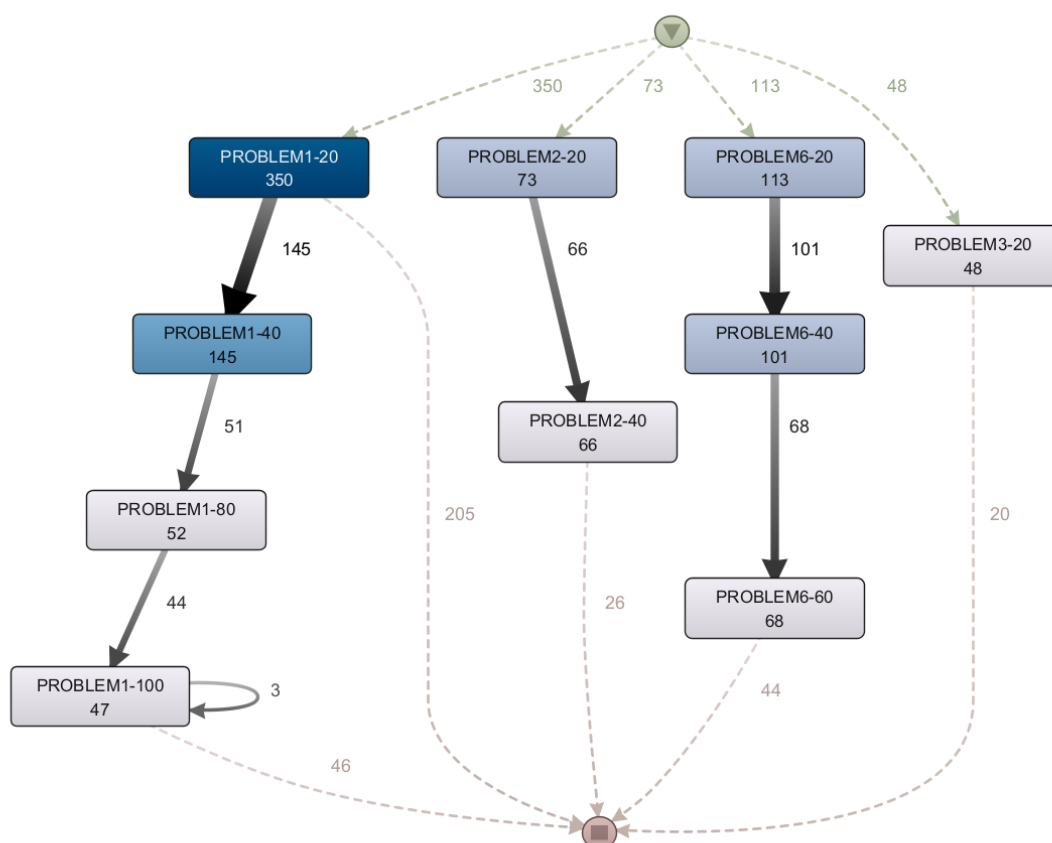


Figura 5.24: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “baja” del curso académico 1617.

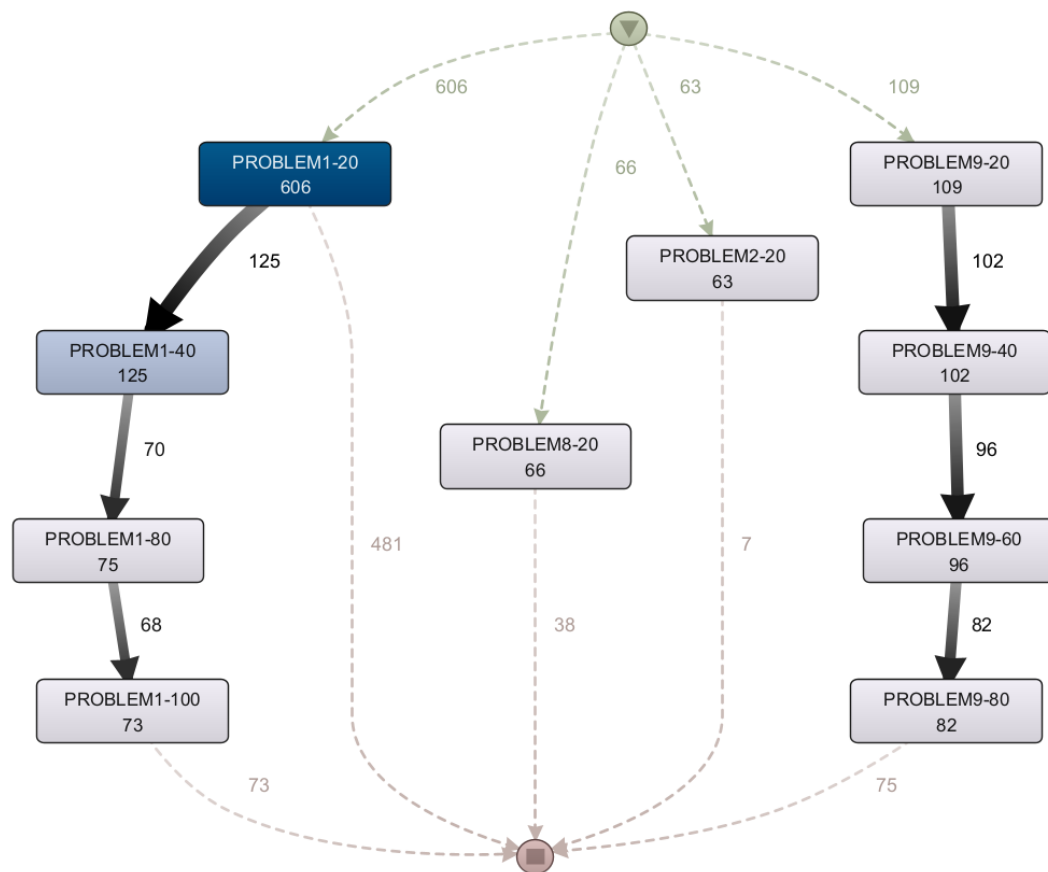


Figura 5.25: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “baja” del curso académico 1718.

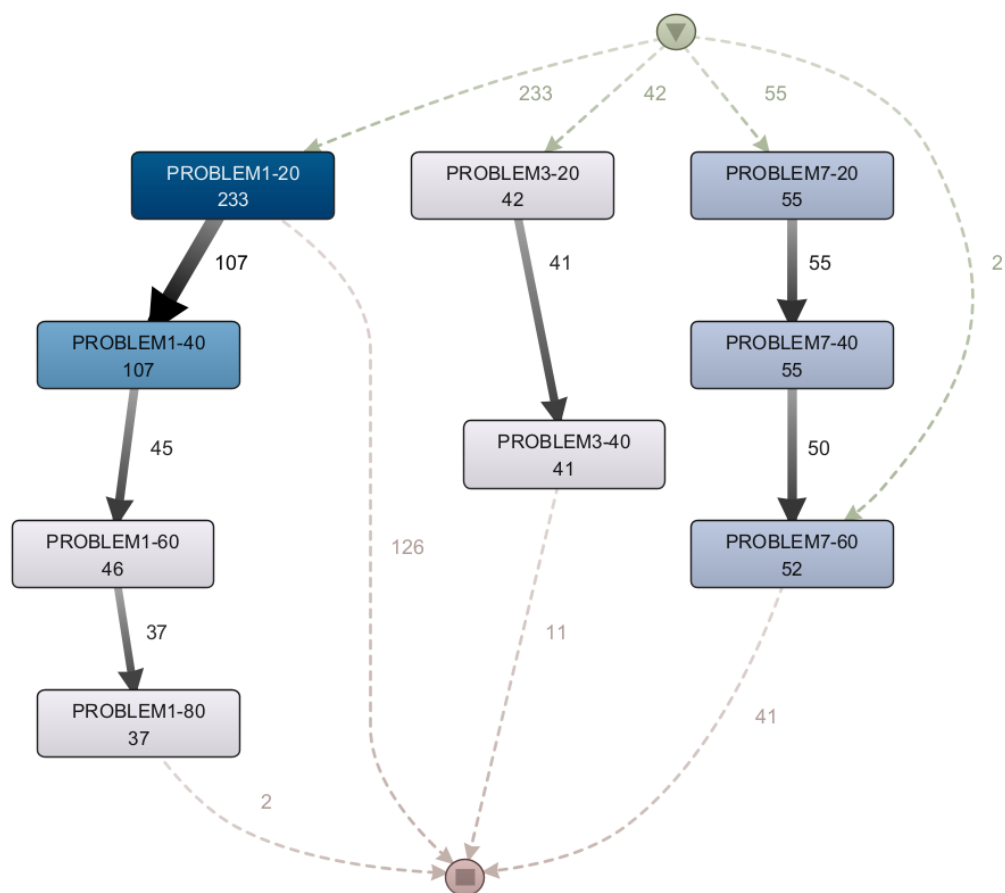


Figura 5.26: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación "baja" del curso académico 1920.

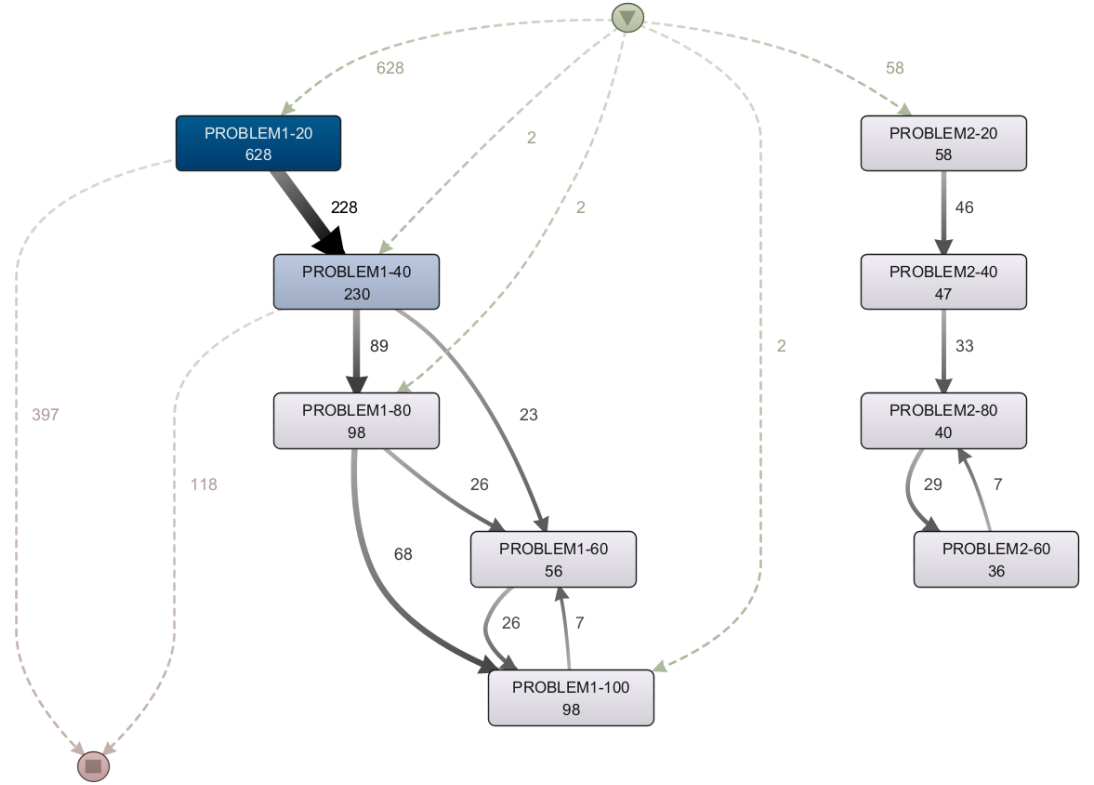


Figura 5.27: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “media-baja” del curso académico 1516.

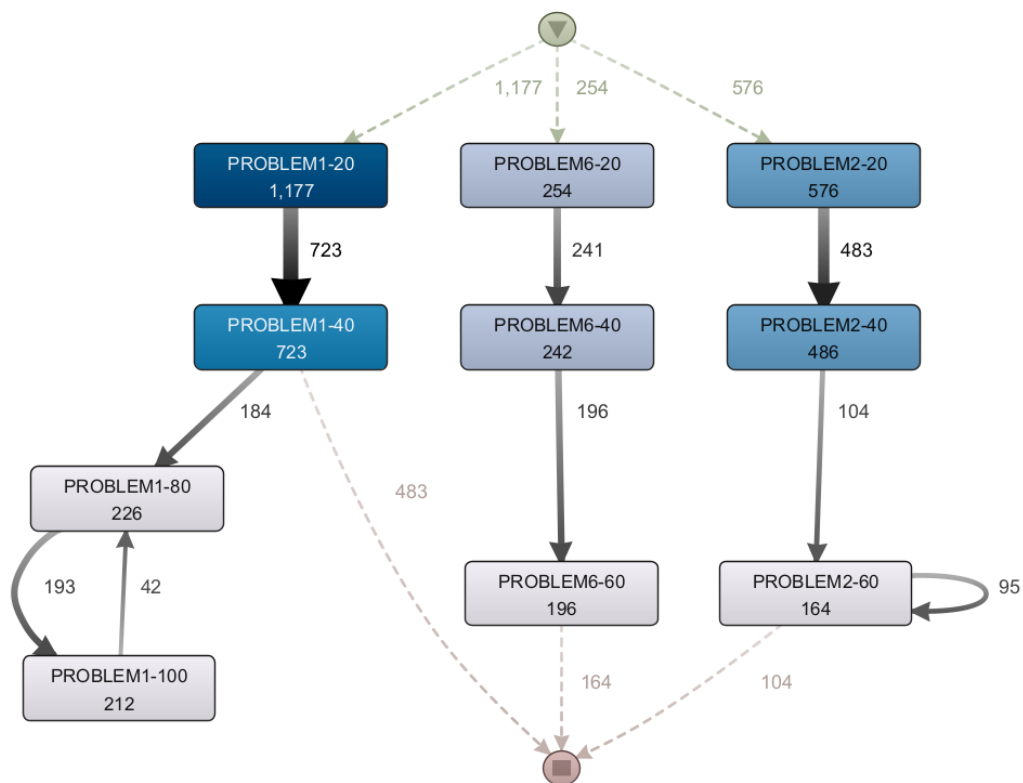


Figura 5.28: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “*media-baja*” del curso académico 1718.

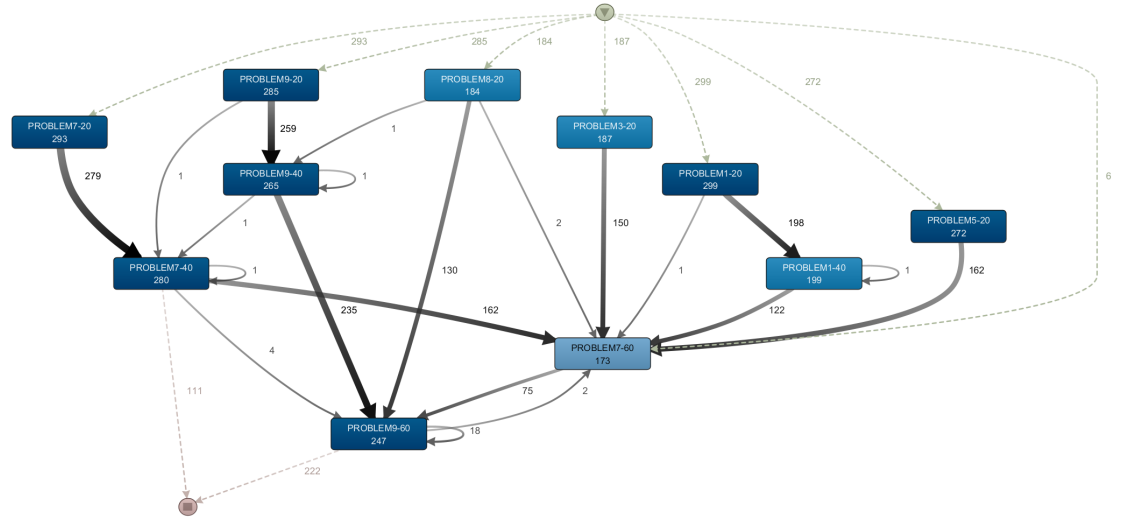


Figura 5.29: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “media-baja” del curso académico 1920.

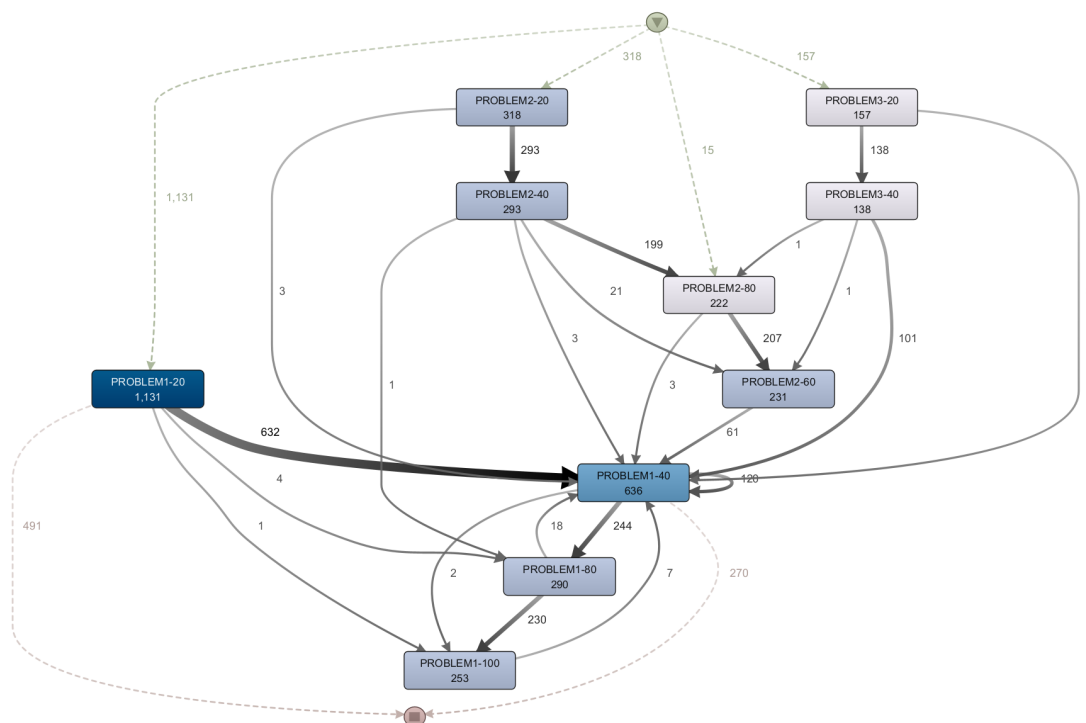


Figura 5.30: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación "media-alta" del curso académico 1516.

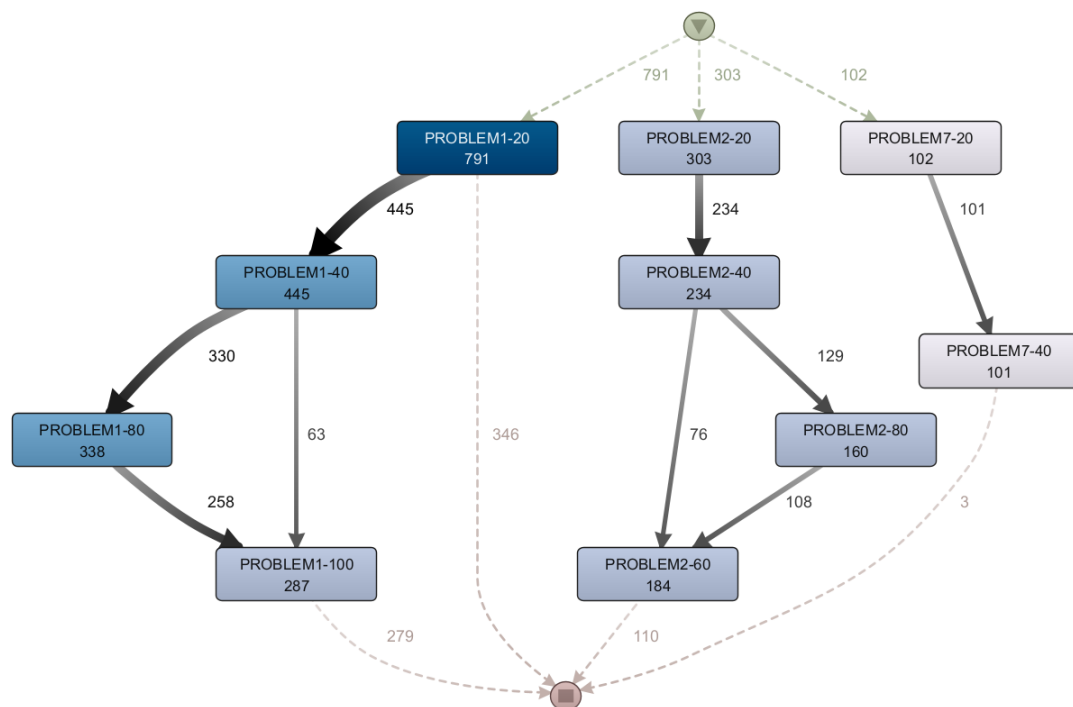


Figura 5.31: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “*media-alta*” del curso académico 1617.

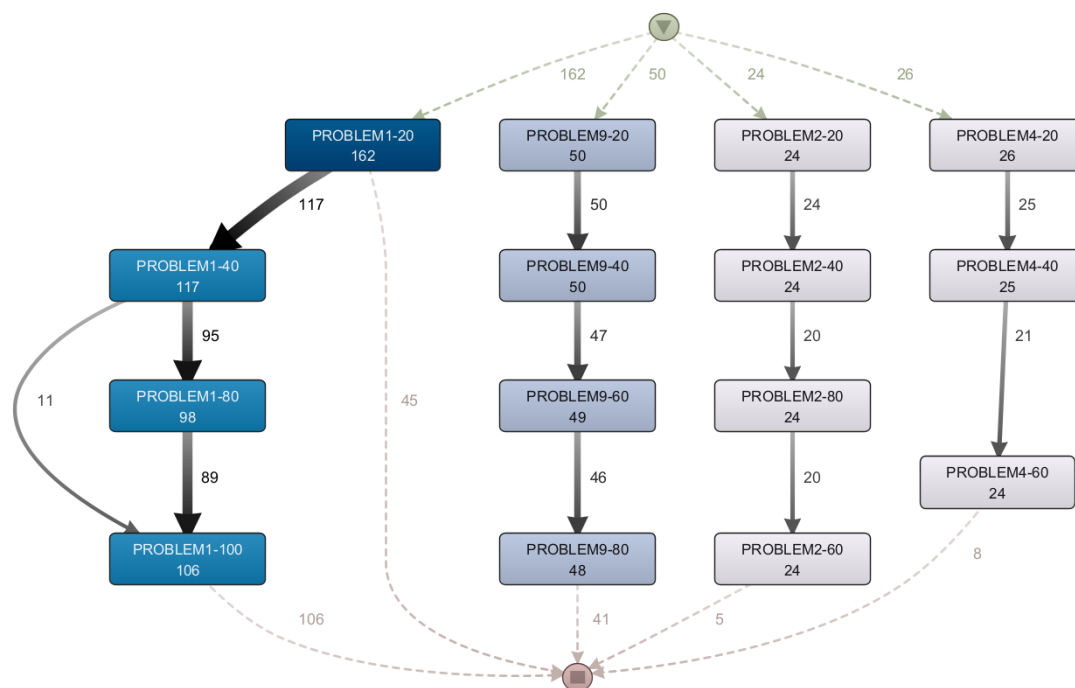


Figura 5.32: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “*media-alta*” del curso académico 1718.

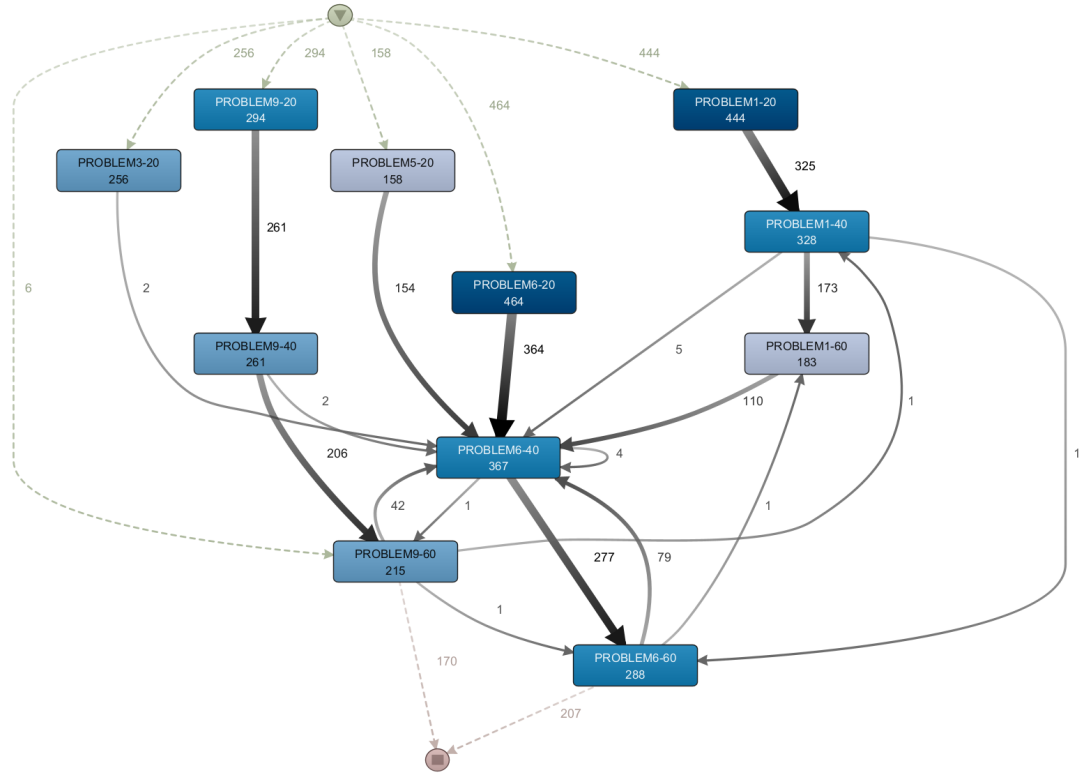


Figura 5.33: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “media-alta” del curso académico 1920.

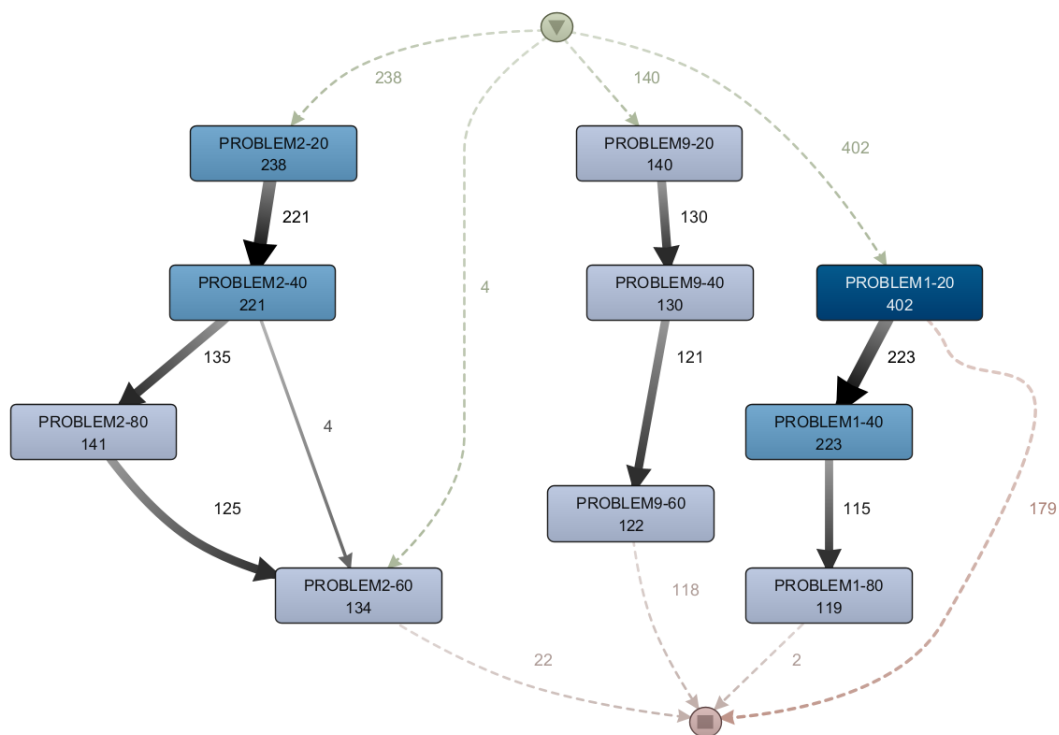


Figura 5.34: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación "alta" del curso académico 1516.

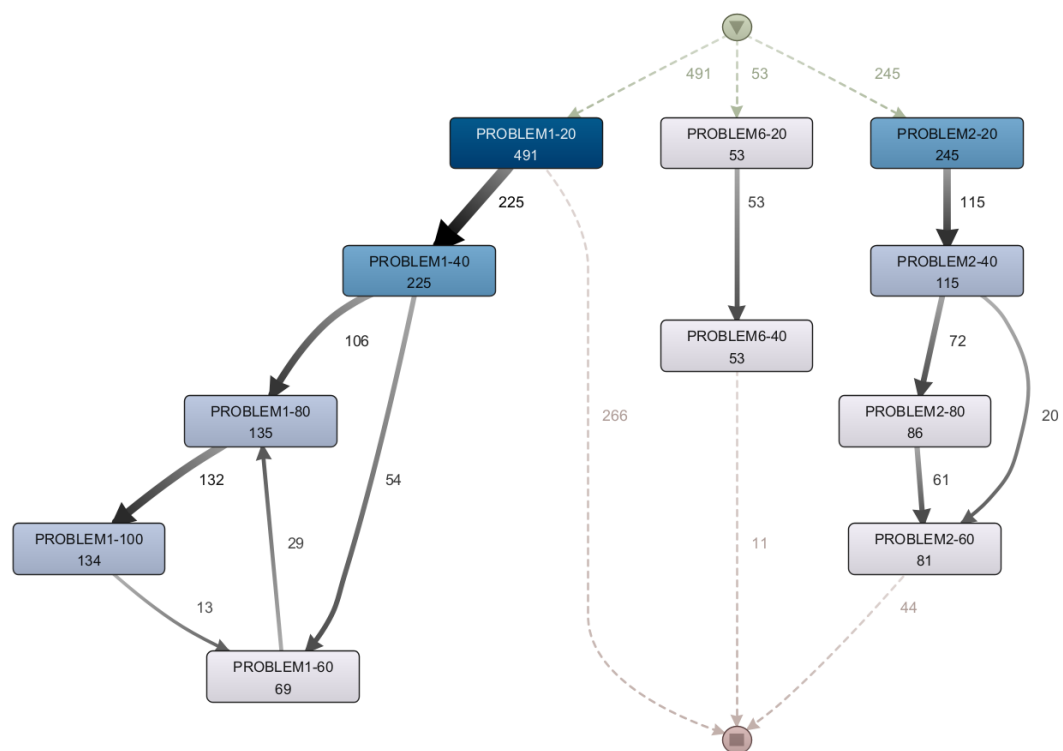


Figura 5.35: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “alta” del curso académico 1617.

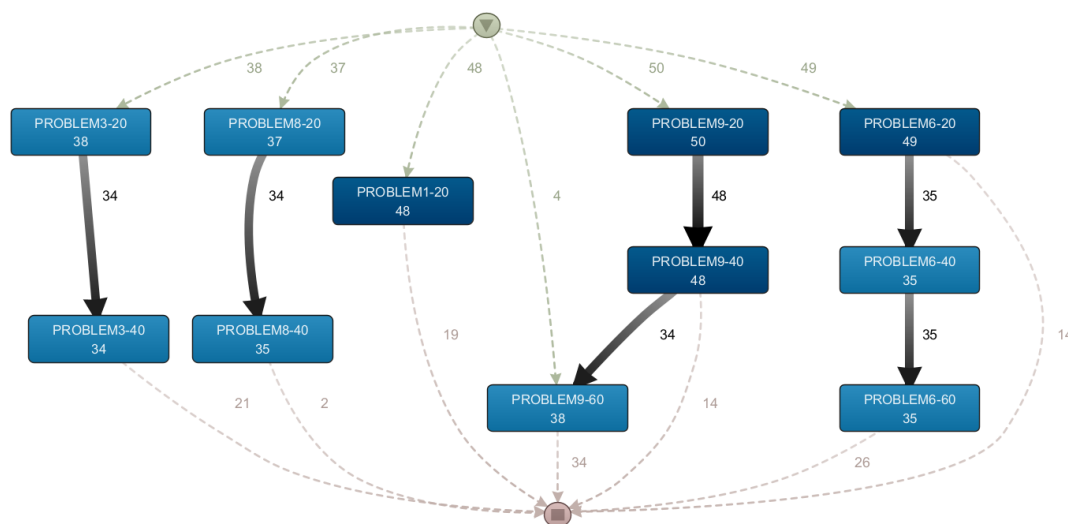


Figura 5.36: Extracción de procesos del dataset integrado por las acciones compuestas de los grupos con calificación “alta” del curso académico 1920.

Como podemos ver en las Figuras 5.34, 5.35 y 5.36, el problema 2 cobra importancia así como otros terceros problemas que varían en función del año:

- El problema 9 es el tercero más frecuentado en el curso 1516.
- El problema 6 ocupa el tercer puesto en el año académico 1617.
- El año 1920 presenta diferencias. El nodo más visitado es el PROBLEM9-20 (frecuencia 50), seguido de los nodos PROBLEM6-20, PROBLEM1-20, PROBLEM3-20 y PROBLEM3-20 (con frecuencias 49, 48, 38 y 37 respectivamente).

Por último, cabe descartar que parece que hay una tendencia a resolver los problemas de manera más secuencial en los procesos extraídos con los grupos de calificaciones más altas. Esto puede observarse claramente si se siguen todas las Figuras de este apartado correspondientes al curso académico 1920 (Figuras 5.26, 5.29, 5.33 y 5.36).

5.3. Análisis de los procesos con segmentación

5.3.1. Arquitectura del software

5.3.2. Ejecución del software

Para la correcta ejecución del software se han de seguir las siguientes indicaciones:

- En primer lugar, se ha de ejecutar el makefile para compilar el código C++ que se encuentra en la carpeta `/code/C++/DAG`.
- A continuación, para obtener los resultados a partir de los diferentes archivos .csv que obtuvimos en esta iteración y en iteración anterior se deberá ejecutar el script que se encuentra en ese mismo directorio.

5.4. Análisis de los motivos en los cambios de comportamiento

5.5. Análisis de si hay cambios inducidos por los éxitos de los demás grupos

